

**INSTITUTO POLITÉCNICO DO PORTO**  
**INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO**  
**DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA**

**Rede de Monitorização Ambiental em Tempo-  
Real com Sistemas Móveis**

**Pedro Manuel Sousa Guimarães**

**(Licenciado)**

**Dissertação para Obtenção do Grau de Mestre em  
Engenharia Electrotécnica - Ramo de Sistemas Autónomos**

**Orientador: Eduardo Alexandre Pereira da Silva (Doutor)**

**Porto, Dezembro de 2009**



## Resumo

A compreensão das interações entre os oceanos, a linha de costa, a qualidade do ar e as florestas só será possível através do registo e análise de informação geo-temporalmente referenciada. Mas a monitorização de grandes áreas apresenta o problema da cobertura espacial e temporal, e os custos nela envolvidos pela impossibilidade de disseminar a quantidade de estações de monitorização necessários à compreensão do fenómeno.

É necessário então definir metodologias de colocação de sensores e recolha de informação de forma robusta, económica e temporalmente útil.

Nesta dissertação apresentamos uma estratégia de monitorização ambiental, para meios hídricos, (ou de grande dimensão) que baseada em sistemas móveis e alguns princípios da geoestatística, fornece uma ferramenta de monitorização mais económica, sem prejuízo da qualidade de informação.

Os modelos usados na geoestatística assentam na ideia de que medidas mais próximas tendem a serem mais parecidas do que valores observados em locais distantes e fornece métodos para quantificar esta correlação espacial e incorporá-la na estimação.

Os resultados obtidos sustentam a convicção do uso de veículos móveis em redes de sensores e que contribuímos para responder à seguinte questão “Qual a técnica que nos permite com poucos sensores monitorizar grandes áreas?”. A solução passará por modelos de estimação de grandezas utilizados na geoestatística associados a sistemas móveis.

**Palavras-chave:** Redes de Sensores, Estimação Espacial, *Krigagem* ordinária

Esta página foi intencionalmente deixada em branco.

## Abstract

Understanding the interactions between the oceans, the coastline, air quality and forests will only be possible through registration and analysis of geo-temporally referenced. But the monitoring of large areas presents the problem of spatial and temporal coverage and the costs involved it was impossible to spread the number of monitoring stations needed to understand the phenomenon.

It is then necessary to define methodologies for placing sensors and collecting information in a robust, economic and temporally useful.

In this thesis we present a strategy for environmental monitoring for water resources (or large) that based on the move and some principles of geostatistics provides a tool for monitoring more economical, without loss of quality of information.

The models used in geostatistics are based on the idea that measures more closely tend to be more similar than those observed in remote locations and provides methods to quantify this spatial correlation and incorporate it in the estimation.

The results support the belief of the use of mobile vehicles in sensor networks and contribute to answer the question "What is the technique that allows us with a few sensors to monitor large areas?". The solution is the estimation models used in geostatistics quantities associated with mobile systems.

**Keywords:** Networked Sensors, Spatial Estimation, Ordinary Kriging

Esta página foi intencionalmente deixada em branco.

## Agradecimentos

Esta tese é o culminar de um objectivo académico que me propus e que não seria possível sem a ajuda de um número considerável de pessoas.

Ao meu orientador, Doutor Eduardo Alexandre Pereira da Silva pela confiança, apoio, paciência e amizade.

Ao Professor Joaquim Eduardo de Sousa Góis do Departamento de Engenharia de Minas da Faculdade de Engenharia da Universidade do Porto, pela ajuda preciosa que me deu na área da geoestatística.

A todos os colegas e membros do Laboratório de Sistemas Autónomos (LSA), principalmente ao Eng.º José Miguel Almeida, Eng.º Alfredo Martins e Eng.º Luís Lima pela ajuda e contribuição que deram a esta tese.

À Doutora Cristina Ribeiro, Dr.<sup>a</sup> Natércia Lima, Eng.º Filipe Pires de Moraes e Eng.º Frederico Jacob agradeço a honra de tê-los como amigos e pela dedicação de todos os dias.

Aos meus pais pelo estímulo, serenidade e apoio incondicional desde a primeira hora e pela sua compreensão, amor e carinho.

Finalmente agradeço a todos os que directa ou indirectamente colaboraram para esta tese nomeadamente os colegas do Departamento de Física e do ISEP, alunos e a todos os meus amigos.

Pedro Manuel Sousa Guimarães

Esta página foi intencionalmente deixada em branco.



## Lista de Símbolos

$2\gamma(\mathbf{h})$	Variograma;
$\gamma(\mathbf{h})$	Semivariograma ;
$\mathbf{h}$	Vector Distância (modulo e direção) entre pares de observação;
$N(\mathbf{h})$	Número de pares de valores medidos, separados por uma distância $\mathbf{h}$ ;
$Z(\mathbf{x}_i)$	Valor recolhido no ponto $\mathbf{x}_i$
$Z(\mathbf{x}_i + \mathbf{h})$	Valor recolhido no ponto $\mathbf{x}_i + \mathbf{h}$

Esta página foi intencionalmente deixada em branco.

## Acrónimos

ESN	<i>Environmental Sensor Network</i>
LSA	Laboratório de Sistemas Autónomos
GDI	<i>Great Duck Island</i>
ORCA	<i>Oceanic Remote Chemical/optical Analyzer</i>
NOAA	<i>National Oceanic &amp; Atmospheric Administration</i>
GLERL	<i>Great Lakes Environmental Research Laboratory</i>
BMELib	<i>Bayesian Maximum Entropy Library</i>
US-EPA	<i>United States Environmental Protection Agency</i>
CTD	<i>Conductivity, Temperature, and Depth</i>
ARM	<i>Advanced Risk Machine</i>
GTK	<i>GIMP Tool Kit</i>
LINCE	<i>Land Intelligent Cooperative Explorer</i>
RS	Rede de Sensores
ALERT	<i>Automated Local Evaluation in Real Time</i>
VR	Variáveis Regionalizadas
BW	Largura de Banda
ROV	<i>Remotely operated vehicle</i>
UAV	<i>Unmanned Aerial Vehicle</i>
GPS	<i>Global Positioning System</i>
RFID	<i>Radio-Frequency IDentification</i>

Esta página foi intencionalmente deixada em branco.

## Índice

Resumo .....	3
Abstract.....	5
Agradecimentos .....	7
Lista de Símbolos.....	9
Acrónimos.....	11
1 Introdução.....	19
1.1 Âmbito da Dissertação .....	19
1.2 Motivação e enquadramento .....	22
1.3 Objectivos.....	23
1.4 Organização da Dissertação .....	23
2 Redes de Sensores .....	25
2.1.1 Tarefas Típicas de uma Rede de Sensores .....	26
2.1.2 Componentes de uma rede de sensores .....	27
2.1.3 Caracterização de uma Rede de Sensores .....	30
2.1.4 Redes de Sensores Vs Redes Tradicionais de Dados .....	31
2.2 Redes de Sensores Ambientais ( <i>ESN</i> ).....	32
2.2.1 Exemplos de redes de sensores ambientais .....	34
2.2.2 Redes de larga escala unifuncionais.....	34
2.2.3 Redes de sensores locais multifunção .....	34
2.2.4 Redes de biosensores.....	34
2.2.5 Redes de Sensores Heterogénea .....	35
2.3 Redes de sensores ambientais- Cenários de aplicação .....	35
2.3.1 <i>Great Duck Island (GDI) System</i> .....	35
2.3.2 <i>PODS – A remote ecological micro-sensor network</i> .....	36
2.3.3 Projecto <i>CORIE</i> .....	37
2.3.4 Projecto <i>ORCA</i> .....	38
2.3.5 <i>ALERT - Automated Local Evaluation in Real Time</i> .....	39
2.3.6 <i>James Reserve Extensible Sensing System</i> .....	40
2.3.7 <i>Floodnet</i> .....	40
3 Estimação espacial .....	43
3.1 Introdução.....	43
3.2 Geoestatística e suas aplicações .....	44

3.3	Representação geométricas de dados .....	45
3.3.1	Tipo e natureza de dados em análise espacial .....	45
3.3.2	Representação computacional dos dados geográficos .....	46
3.3.3	Tipo de variáveis ou Natureza de dados.....	46
3.3.4	Análise Estatística .....	47
3.4	Teoria das variáveis regionalizadas.....	50
3.4.1	Análise Geoestatística .....	50
3.4.2	Conceitos Básicos em análise espacial.....	51
3.4.3	Dependência espacial e Autocorrelação espacial.....	52
3.4.4	Características das variáveis regionalizadas .....	53
3.4.5	Variáveis regionalizadas .....	56
3.4.6	Variografia .....	60
3.5	Estimação geostatística .....	69
3.5.1	Introdução.....	69
3.5.2	Estimador Linear Geoestatístico: <i>Krigagem</i> .....	73
3.5.3	Sistema de equações de <i>krigagem</i> .....	74
3.5.4	Exemplo Prático .....	76
4	Sistema Desenvolvido .....	79
4.1	Solução para um veículo autónomo de superfície.....	79
4.1.1	Grandezas físicas.....	80
4.1.2	CTD .....	83
4.1.3	GPS.....	84
4.1.4	Comunicação .....	85
4.2	Amostragem .....	86
4.3	Metodologia .....	87
4.4	Desenvolvimento.....	88
4.4.1	Plataforma <i>Web</i> .....	93
5	Resultados Obtidos.....	97
5.1	Análise exploratória de dados .....	98
5.2	Parametrização .....	99
5.3	Estimação e Geração de Mapas.....	99
5.4	Avaliação de Erros .....	101
6	Conclusões e Trabalhos Futuros .....	103
7	Bibliografia.....	107

## Lista de Tabelas

Tabela 1- Série de amostras de uma grandeza .....	54
Tabela 2- Estatísticas medidas para as duas séries de números .....	54
Tabela 3- Variâncias espaciais para as séries A e B, determinadas até quatro intervalos de amostragem .....	55
Tabela 4: Valores da função semivariograma calculados para as direcções Este- Oeste e Norte – Sul e semivariogramas experimentais [40] .....	65
Tabela 5- Modelos teóricos de ajuste .....	66
Tabela 6: Resumo estatístico dos dados das medidas de temperatura .....	98
Tabela 7 – Parâmetros do modelo ajustado para o semivariograma da temperatura .....	99
Tabela 8- Sensor virtual 1 .....	102
Tabela 9- Sensor virtual 2 .....	102
Tabela 10- Sensor virtual 3 .....	102

## Lista de Figuras

Figura 1- Imagem de uma linha de costa [49].....	19
Figura 2 – Projecto LINCE .....	22
Figura 3 – Projecto FALCOS.....	22
Figura 4 – Projecto ROAZ .....	22
Figura 5: Exemplo de uma Rede de Sensores .....	25
Figura 6 - Classificação de um objecto usando RS .....	27
Figura 7- Monitorização\Mapeamento de um evento .....	27
Figura 8 – Diagrama esquemático representando as partes de uma rede de sensores ambientais [25] .....	28
Figura 9 - Espectro de dispositivos de sensores [25] .....	34
Figura 10- Localização espacial das estações .....	37
Figura 11- Sistema móvel marítimo.....	37
Figura 13- Componentes da variação espacial de uma variável regionalizada [45] .....	56
Figura 14 – Aspecto Aleatório de uma variável regionalizada [45] .....	57
Figura 15 – Hipótese de estacionaridade [45].....	58
Figura 16 - Hipótese de estacionaridade de 2ª ordem [45] .....	59

Figura 17: Dependência entre duas variáveis regionalizadas.....	61
Figura 18- Semivariograma experimental [38] .....	63
Figura 19 – Conjunto de amostras regularmente espaçadas[50].....	63
Figura 20 - Parâmetros de um semivariograma para amostras irregularmente espaçadas [45] 65	
Figura 21 – Fenómeno isotrópico [40].....	68
Figura 22 - Fenómeno anisotrópico [40].....	68
Figura 23 - Anisotropia geométrica .....	69
Figura 24- Anisotropia zonal [40].....	69
Figura 25 - Anisotropia combinada.....	69
Figura 26 - Pontos de amostragem numa área de estudo [51] .....	70
Figura 27 - Pontos de amostragem e de estimativa numa área de estudo [51] .....	71
Figura 28 – Localização espacial de pontos de amostragem [40].....	76
Figura 29 – Inter-relações entre amostras e o ponto a ser estimado para cálculo das equações de krigagem. [40] .....	77
Figura 30 –Modelo de variograma para os dados da figura 28.....	77
Figura 31 - CTD .....	80
Figura 32 - Veículo autónomo de superfícies - ROAZ .....	80
Figura 33 - Sensor de pressão .....	82
Figura 34 - Receptor de navegação <i>Superstar II</i> .....	84
Figura 35 - Interface gráfico com o valor das grandezas .....	86
Figura 36 - Localização das estações do Programa <i>Costwatch</i> .....	87
Figura 37 - Localização das estações ao longo do <i>Lake Erie</i> .....	88
Figura 38: Exemplo do ficheiro XML com os pontos amostrados e respectivos valores da temperatura.....	90
Figura 39- Exemplo de um variograma experimental.....	91
Figura 40: Exemplo de mapa gerado.....	91
Figura 41 – Outro exemplo de um mapa gerado.....	91
Figura 42: Mapa gerado com vista à sua utilização na <i>Web</i> .....	92
Figura 43 - Representação dos pontos recolhidos por estações fixas e veículos móveis.....	93
Figura 44 - Representação dos pontos recolhidos pelo veículo móvel e respectivo percurso. 94	
Figura 45 - Representação dos pontos recolhidos pelo veículo móvel e respectivo percurso. 94	
Figura 46 - Aplicação <i>Google Earth</i> com a sobreposição do mapa gerado. ....	95
Figura 47 - Localização espacial das estações fixas. ....	97
Figura 48 – Trajecto efectuado por um veículo móvel. ....	97



Figura 49 - Localização espacial de todos os pontos recolhidos.....	97
Figura 51 – Respectiva distribuição espacial .....	98
Figura 52 –Semivariograma experimental Omnidireccional da temperatura .....	99
Figura 53 - Mapa da distribuição da temperatura na zona dos Grandes Lagos ( <i>Lake Erie</i> )..	100
Figura 54- Mapa da distribuição da temperatura às 00:00 .....	100
Figura 55 - Mapa da distribuição da temperatura às 04:00 .....	100
Figura 56 - Mapa da distribuição da temperatura às 08:00 .....	101
Figura 57 - Mapa da distribuição da temperatura às 12:00 .....	101
Figura 58 - Mapa da distribuição da temperatura às 16:00 .....	101
Figura 59 - Mapa da distribuição da temperatura às 20:00 .....	101

Esta página foi intencionalmente deixada em branco.

# 1 Introdução

## 1.1 Âmbito da Dissertação

Será fundamental para o sucesso da humanidade num futuro próximo que as tomadas de decisão com forte impacto social e económico sejam baseadas em informação tempo-real e predições com elevado índice de confiabilidade.

Só assim o crescimento económico será viável e compatível com uma saúde pública robusta, actividades de lazer igualmente disseminadas pela população, redução dos impactos de eventos severos como "*tsunamis*" e adaptação às mudanças de clima. Neste contexto surge como fundamental as relações que se estabelecem entre os oceanos, a linha de costa, a qualidade do ar e as florestas.

A compreensão destas interações e o conhecimento da evolução do estado da terra só será possível através do registo e análise de informação geo-temporalmente referenciada.

A caracterização do estado dos recursos naturais tal como florestas, recursos hídricos e ecológicos depende sobretudo da dispersão [18] espacial e espaço-temporal das grandezas físicas tais como temperatura, humidade, radiação luminosa, etc.

É fundamental definir estratégias de colocação de sensores e recolha de informação de forma robusta, económica e temporalmente útil a grande escala e cobrindo grandes áreas, por forma a instrumentalizar o meio físico, como por exemplo em ambientes estuarinos, costeiros ou florestais, bem como a caracterizar os factores que influenciam na sua distribuição espacial e na sua variação temporal.



Figura 1- Imagem de uma linha de costa [49]

É necessário o aparecimento de gerações de sensores de baixo valor, biodegradáveis que possam ser colocados em larga escala com interfaces adaptáveis aos sucessivos desenvolvimentos tecnológicos.

A vulgarização destes dispositivos de sensorização levará à criação e desenvolvimento de redes de sensores que nada mais são do que um conjunto de sistemas autónomos que para além de capacidades de medição de grandezas físicas possuem capacidades de processamento, gestão de energia e comunicação. [19]

Adicionalmente, será necessário o aumento e banalização da utilização de plataformas autónomas (veículos submarinos, superfície e aéreos) capazes de transportar e implementar estratégias de observação que dependam dos fenómenos em causa e da localização geográfica.

Existem actualmente muito exemplos de redes de monitorização ambiental com aplicações que vão desde a monitorização de *habitats* [2] [6], vigilância de ambientes inhóspitos, controlo de parâmetros hidrológicos, químicos e biológicos da água [3] bem como na previsão e prevenção da expansão de catástrofes naturais. [41]

No âmbito destas redes, é de salientar a *Reserve Extensible Sensing System* [2] e o Projecto *CORIE* [3] por integrarem sistemas móveis como elementos fundamentais no registo da informação. A introdução de mobilidade visa sobretudo a recolha de imagem em locais remotos e/ou aquisição e registo de informação em zonas onde a rede apresente lacunas.

Compreender a distribuição espacial de dados oriundos de fenómenos ocorridos no espaço constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento [19]. No entanto quando não se possui um modelo que descreva um determinado fenómeno físico não conseguimos saber qual ou quais os valores nas posições espaciais não amostradas.

Os modelos usados na geoestatística assentam na ideia de que medidas mais próximas tendem a serem mais parecidas do que valores observados em locais distantes e fornece métodos para quantificar esta correlação espacial e incorporá-la na estimação e na inferência, ao contrário da estatística clássica que ignora a dependência entre amostras.

Isto possibilita criar um novo método de monitorização “virtual”, onde mesmo em que um determinado ponto não possua um sensor, a proximidade de outros sensores e análise dos dados permitem estimar a sua leitura [30].

O problema de encontrar o melhor estimador foi resolvido por *G. Matheron* e que em homenagem a *Daniel G. Krige* deu-lhe o nome de *Krigagem*.

A diferença entre a *krigagem* e outros métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras pois usa informações a partir do semivariograma

para encontrar os pesos óptimos a serem associados às amostras que irão estimar um ponto, um área ou um bloco.

Os problemas de natureza ambiental são demasiado complexos para serem explicados através de modelos determinísticos (uma das excepções ocorre na modelização da dinâmica das plumas de contaminação do meio). Pelo facto de não se poder repetir a amostra ou observação num espaço e no tempo, os fenómenos relacionados com as ciências da terra e do ambiente, possuem um grau de incerteza elevado ligado ao seu conhecimento [18].

O estudo de sistemas físicos como oceanos, rios ou lagos, caracterizando as suas propriedades, compreendendo os processos que neles ocorrem e as interações com os sistemas vizinhos, a previsão do seu comportamento são matéria de estudo de diferentes ciências.

Mas a observação dos meios hídricos apresenta dificuldades acrescidas relativamente à de outros sistemas físicos por se tratar por vezes de um meio extenso (caso do oceano), de difícil acesso, que é praticamente opaco à radiação electromagnética, e que obriga os instrumentos de medição a suportar fortes pressões e uma corrosão intensa.

Os métodos de observação utilizados têm evoluído ao longo do tempo em função do desenvolvimento tecnológico que se vai atingindo. O rápido progresso da electrónica veio modificar drasticamente o modo de observar o oceano, permitindo descobrir muitos aspectos até aí ignorados e a existência de uma enorme diversidade de escalas espaciais e temporais. [43].

A monitorização destas grandes áreas apresenta o problema da cobertura espacial e temporal, e aos custos nela envolvidos pela impossibilidade de disseminar a quantidade de estações de monitorização necessários à compreensão do fenómeno.

Nesta dissertação apresentamos uma estratégia de monitorização ambiental, para meios hídricos (ou de grande dimensão), que baseada em sistemas móveis e alguns princípios da geoestatística, fornece uma ferramenta de monitorização mais económica, sem prejuízo da qualidade de informação. Complementarmente apresentamos o desenvolvimento de um sensor de monitorização de Condutividade e Temperatura, para equipar um sistema de móvel de monitorização de ambientes hídricos.

É apresentada uma alternativa às actuais redes de monitorização remota com o objectivo de que o utilizador consiga estimar uma grandeza em áreas não amostradas, suportadas em técnicas utilizadas na geoestatística, a partir de observações provenientes de elementos fixos da rede. Os elementos móveis circulam a área e fornecem ao modelo utilizado valores de zonas não amostradas que em comparação com os valores estimados permite

calibrar o modelo através do erro entre valores estimados e valores reais em zonas não amostradas no tempo.

Também podem ser observados nesta dissertação resultados experimentais que permitem avaliar preliminarmente a estratégia de monitorização por nós sugerida, com particular destaque para o exemplo de utilização que testamos com informação proveniente do sistema de monitorização dos Grandes Lagos desenvolvida pelo “*Great Lakes Environmental Research Laboratory*”.

Dos resultados obtidos podemos afirmar que definimos uma estratégia que nos permite obter a mais informação a menor custo. E que contribuímos para responder à seguinte questão “Qual a técnica que nos permite com poucos sensores monitorizar grandes áreas?”. A solução passará por modelos de estimação de grandezas utilizados na geoestatística associados a sistemas móveis.

## 1.2 Motivação e enquadramento

No Laboratório de Sistemas Autónomos (LSA-ISEP) e enquadrada na sua actividade de I&D existem diversos projectos de veículos autónomos com a possibilidade de operar em diferentes meios.



Figura 2 – Projecto LINCE



Figura 3 – Projecto FALCOS



Figura 4 – Projecto ROAZ

Existe também um esforço para desenvolver sistemas sensoriais inovadores/dedicados que permitam um aumento das capacidades de percepção dos veículos desenvolvidos com aplicabilidade nos meios oceânico, aéreo e terrestre.

O desenvolvimento das aplicações orientadores de cenários que vão desde a monitorização ambiental ou vigilância marítima (entre outros) promovem diferentes desafios para os sistemas promovem diferentes desafios para os sistemas de percepção.

Este quadro carece de um aprofundamento do estudo, do controlo activo e da exploração dos fenómenos que se desenvolvem. No entanto, os custos económicos associados ao incremento da actividade de observação baseada nas tecnologias actualmente usadas tornam-se in comportáveis.

Com o objectivo de ter a máxima informação a um custo mínimo, pretende-se verificar o benefício do uso de veículos móveis, em sistemas distribuídos, para a estimação de parâmetros físicos recorrendo à modelização geoestatística, ou seja no uso de sensores virtuais

### **1.3 Objectivos**

É objectivo desta dissertação promover o desenvolvimento de uma solução sistémica que permita de forma simples e a custos reduzidos realizar missões de detecção e de observação de fenómenos em meios naturais principalmente nos aquáticos, ou nas estruturas confinantes.

Pretende-se a concepção de uma estrutura que permita o registo, a difusão e o desenvolvimento em tempo-real de modelos de caracterização de fenómenos naturais.

Assim são objectivos:

- Desenvolvimento de uma solução para um veículo autónomo de superfície que permita a recolha e a georeferenciação de dados.
- A definição de uma estratégia de observação que através da utilização de sistemas móveis, em complementaridade de sistemas fixos, permita através de modelos de referência da geoestatística , obter informação caracterizadora do meio.
- Avaliar os resultados experimentais da abordagem metodológica proposta ou seja, como é que os modelos geoestatísticos, com calibrações frequentes com veículos móveis, podem ou não apresentar melhores resultados.

### **1.4 Organização da Dissertação**

No capítulo 1 é feita a contextualização desta dissertação assim como definida a sua motivação, objectivos e contribuição desta tese.

O capítulo 2 resume o estado de arte no que diz respeito a redes de sensores. São descritas as aplicações das redes de sensores a nível ambiental e feita uma descrição dos projectos mais importantes desenvolvidos nesta área.

No capítulo 3 é feita a apresentação dos principais conceitos teóricos da geoestatística necessários ao desenvolvimento deste trabalho. Inicia-se com a definição de conceitos de estimação espacial, para de seguida fazer a contextualização das variáveis regionais, terminando o capítulo com a apresentação dos métodos de estimação de *krigagem*.

No capítulo 4 é descrita a implementação do sistema e de cada componente que servirá para integração num veículo móvel aquático de superfície. São igualmente apresentados principais módulos de *software* desenvolvido com vista à estimação espacial.

No capítulo 5 são apresentados os resultados dos testes efectuados com os dados adquiridos.

As conclusões deste trabalho e sugestões para futuros desenvolvimentos são apresentados no capítulo 6.



## 2 Redes de Sensores

Uma rede de sensores (RS) pode-se caracterizar por possuir sistemas autónomos (elementos capazes de adquirir dados do meio envolvente e distribuir esses dados através da rede) com diferentes capacidades de processamento, diferentes modos de transmitir informação e dispositivos com mobilidade distintas, como se mostra na figura 5, no entanto a sua arquitectura depende sempre do seu objectivo principal e da área de aplicação.

As redes de sensores devem possuir alta escalabilidade, de modo a permitir que um grande número de nós possa fazer parte de uma rede. Estes nós devem consumir pouca energia porque uma das ideias básicas de uma rede de sensores sem fio é fazer com que o nó seja um elemento descartável, tornando bastante simples a introdução de sensores em ambientes perigosos ou inóspitos. Devido a este facto, sempre haverá uma constante entrada e saída de nós na rede e por isso as redes de sensores devem ser auto-configuráveis, adaptando-se às novas realidades impostas pelas mudanças na disposição dos sensores. [19]



**Figura 5: Exemplo de uma Rede de Sensores**

As redes de sensores possuem três elementos principais: o sensor, o observador e o fenómeno. [10],[15]

O sensor é o dispositivo que implementa a monitorização física de um fenómeno ambiental e gera relatórios de medidas (através de comunicação sem fio). Um sensor produz uma resposta mensurável a mudanças em condições físicas, tais como temperatura, campo magnético e luz [16]. Os dispositivos de detecção, geralmente, têm características físicas e teóricas diferentes. Muitos modelos de complexidade variada podem ser construídos baseados

na necessidade da aplicação e características dos dispositivos. Na maioria dos modelos de dispositivos sensores a habilidade de detecção diminui com o aumento da distância do sensor ao fenómeno e melhora com o aumento do tempo que o sensor fica exposto para colectar informações [16].

O observador é o utilizador final interessado em obter as informações disseminadas pela rede de sensores em relação a um fenómeno. Ele pode indicar interesses (ou consultas) para a rede e receber respostas destas consultas. Além disso, podem existir, simultaneamente, múltiplos observadores numa rede de sensores.

O fenómeno é a entidade de interesse do observador, que está sendo monitorizada e cuja informação potencialmente será analisada/filtrada pela rede de sensores. Além disso, múltiplos fenómenos podem ser observados concorrentemente numa rede. Numa aplicação, o observador está interessado em monitorizar o comportamento do fenómeno sob algum requisito de desempenho específico (por exemplo, precisão ou atraso). Numa rede de sensores típica, os sensores individuais apresentam amostras de valores locais (medidas) e disseminam informação, quando necessário, para outros sensores e eventualmente para o observador. As medidas realizadas pelos sensores são amostras discretas do fenómeno físico, sujeito a medidas precisas do sensor individual, assim como a localização com respeito ao fenómeno.

### **2.1.1 Tarefas Típicas de uma Rede de Sensores**

Os objectivos de uma rede de sensores, dependem sempre do tipo de aplicação. De seguida são apresentadas algumas delas: [8], [17]

- Determinação do valor de algum parâmetro num dado local. A rede monitoriza constantemente uma ou mais variáveis físicas do ambiente em diferentes locais.
- Determinação da ocorrência de eventos de interesse e estimar valores de parâmetros em função do evento detectado. Na rede os dispositivos são programados para a sensorização de eventos, isto é, padrões que modificam a área de interesse, como por exemplo detecção de movimento.
- Classificação de objectos. Logo após se determinar um evento, pode ser necessário identificar o evento em questão, como por exemplo fogo em áreas florestais (figura 6) ou identificar se o objecto é uma mota ou carro e se ultrapassa uma determinada velocidade.

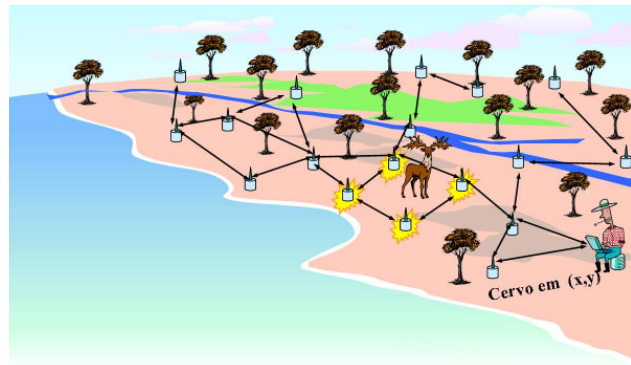


Figura 6 - Classificação de um objecto usando RS

- Monitorização/Mapeamento de um objecto: Além de detectar um evento e identificá-lo, a rede pode ter como tarefa o mapeamento da progressão desse evento. (Figura 7)

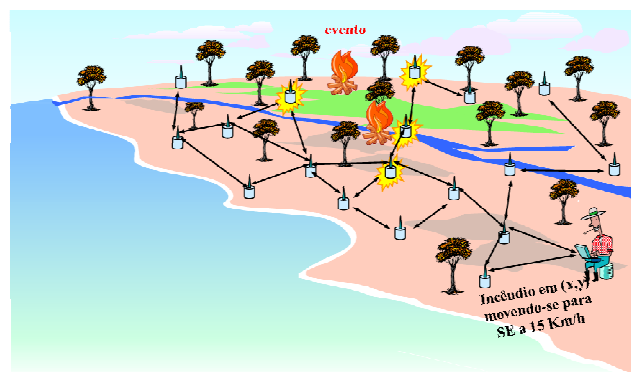


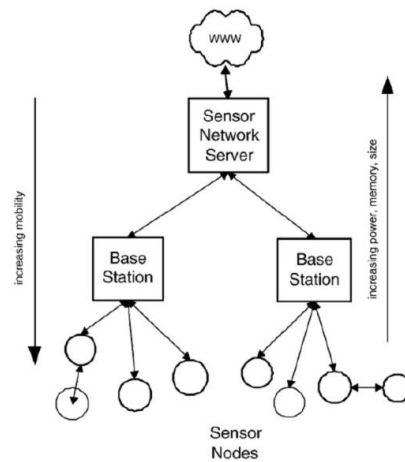
Figura 7- Monitorização\Mapeamento de um evento

### 2.1.2 Componentes de uma rede de sensores

Uma rede de sensores ambiental é constituída por um conjunto de nós sensores, um sistema de comunicações que permite que os dados cheguem a um servidor.

Os nós sensores recolhem dados de uma forma autónoma e uma rede é usada para transmitir dados para uma ou mais estações base que depois irá transmiti-la para um servidor (*Sensor Network Server*).

Os nós sensores podem-se encontrar ligados por cabos existindo muitos projectos que usam esta tecnologia, principalmente as redes subaquáticas, no entanto muitas das aplicações de redes de sensores baseiam-se no *wireless* pois o uso de cabos é muitas vezes impraticável, intrusivo e pode perturbar o meio ambiente a ser monitorizado.



**Figura 8 – Diagrama esquemático representando as partes de uma rede de sensores ambientais [25]**

Tipos diferentes de dados podem ser recolhidos por um nó sensor, como parâmetros ambientais específicos ou mesmo dados GPS diferencial. Os dados podem apresentar diferentes formas: digital ou analógico, espacial ou temporal, alfanumérico ou imagem, fixo ou móvel.

Ao nível do servidor os dados podem ser visualizados dentro de um Sistema de Informação Geográfica em combinação com imagens de satélite ou mapas e publicados através da *Web* para que utilizadores tenham acesso à informação.

Se subirmos na hierarquia do nó sensor para o servidor de dados, vamos assistir igualmente a uma aumento do poder computacional e armazenamento enquanto que a mobilidade dos nós sensores e estações base é maior mas exigem sistemas de localização.

Os nós não necessitam de serem necessariamente estáticos. Para além de poderem ser movidos pelo ambiente, podem ser igualmente movidos através de uma espécie de teleféricos [29], veículos robotizados [27], aéreos [28], micro-submarinos. [26]

É certo de que projectar uma rede de sensores não é uma tarefa trivial pois há a necessidade de entender qual o objectivo que pretende atingir para que todos os requisitos inerentes sejam devidamente avaliados.

Desta forma é necessário a análise de factores e suas implicações no sistema para que possa encontrar o melhor modelo para a rede de sensores que pretendemos implementar. [9] Estes factores podem ser divididos em três categorias [24]. Os primeiros factores a ter em conta são os relacionados com as tecnologias usadas nos nós sensores. Elementos como taxa de transmissão dos nodos, alcance de transmissão e tempo de bateria são aspectos importantes

a ter em conta, no entanto outros como o preço do nó sensor, a sua integridade e a viabilidade de suporte para diferentes sensores são importantes para uma boa implementação.

Factores relacionados com a tecnologia, topologia e arquitectura a ser empregue na implementação da rede são determinados pelas particularidades e características de cada aplicação. Assim a actualização de dados provenientes da monitorização de uma central nuclear terá de ser feita em períodos de tempo muito curtos pois a latência e redundância dos dados precisa de ser maior ao invés de uma aplicação de monitorização ambiental cujos dados podem ser actualizados de 50 em 50 minutos.

Por fim os factores relacionados com a integração do sistema como um todo são de extrema importância nos projectos de redes de sensores. Os dados terão de ser economicamente e inteligentemente transmitidos (idealmente bidireccionalmente) a partir de locais remotos e de alguma forma sejam apresentados aos observadores interessados. Alertas podem ser gerados quando algum problema ocorra ou algum valor seja excedido. Além disso a inserção, troca ou remoção da rede deverá ter o mínimo impacto na operação do sistema.

Projectar e implementar redes de sensores marinhas e aquáticas exige o conhecimento de factores específico deste tipo de redes [13]. O desenvolvimento de observatórios marinhos e costeiros tem sido possível graças a uma série de evoluções tecnológicas, incluindo a miniaturização da electrónica e sensores, ao rápido desenvolvimento da indústria de veículos operados remotamente (ROV), desenvolvimento de conectores subaquáticos fiáveis para ligações eléctricas e ópticas, contínuos avanços nos *housings* subaquáticos e novas abordagens na área da comunicação. No entanto existem factores e componentes críticos que ainda não se encontram adequadamente cobertos por esta tecnologia.

As comunicações são uma questão crítica em observações aquáticas marinhas. Embora sistemas comerciais, em alguns casos, proporcionem excelentes opções para a monitorização costeira (GSM e cobertura de redes sem fios), em observações marinhas mais distantes da costa não têm as mesmas opções comerciais para comunicações terrestres. Os principais requisitos nas comunicações marítimas incluem: longevidade; largura de banda; antenas direccionais, que exigem acompanhamento e estabilização; sistema de armazenamento e transmissão; latência e comunicações bidireccional.

A infra-estrutura para sistemas de observação aquática e marinha inclui necessariamente a manutenção de ROVs e navios, bem como financiamento sustentado de pessoal qualificado. Sensores químicos e biológicos para fornecerem valores com alguma fiabilidade devem ser limpos regularmente e associados à calibração, problema para quase todas as medições

ambientais, tornam a instrumentação uma questão particularmente difícil de investigação e desenvolvimento.

### **2.1.3 Caracterização de uma Rede de Sensores**

A classificação de uma rede de sensores, descrita em [4], é feita para uma rede de sensores sem fios no entanto aplicável a uma rede de sensores genérica.

A aplicação influenciará directamente nas funções exercidas pelos nós, assim como na arquitectura (processador, memória, sensores e fonte de energia), na quantidade de nós que compõem a rede, na distribuição, na escolha dos protocolos da pilha de comunicação, no tipo de dado, no tipo de serviço e consequentemente no tempo de vida dos dispositivos. [4]

Na configuração de uma rede se todos os nós apresentarem a mesma capacidade computacional e de hardware a rede é homogénea, se por ventura a rede possuir uma grande gama de arquitecturas a rede é designada heterogénea.

No que diz respeito à mobilidade a rede pode ser estacionária se possuir nós fixos que permanecem no local onde foram depositados ou móvel caso esses elementos possuam capacidade de mobilidade.

Uma rede pode agrupar uma série de nós (*cluster*) onde a obrigação de recolha, concentração e transmissão de dados esteja a cargo apenas de um elemento (*cluster-head*). Por possuir um nó com estas funções a rede é hierárquica. Numa rede plana todos os nós recolhem e transmitem informações não havendo ninguém destinado para essa tarefa.

O número de nós que se coloca numa determinada área e a maneira como se encontram distribuídos são também parâmetros importantes na caracterização de uma rede. Uma rede pode apresentar uma concentração de nós elevada numa determinada área no entanto nessa área a sua distribuição pode não ser uniforme. De relembrar mais uma vez que a densidade e a distribuição de nós é sempre função do objectivo para a qual a rede foi projectada [4].

A configuração de uma rede põe em evidência aspectos de localização espacial dos nós numa determinada rede mas outros aspectos ligados há a ter em consideração.

A recolha de dados dependendo do objectivo da rede pode ser feita de uma forma contínua ou em intervalos de tempo regulares (sensores de luminosidade encontram-se desligados à noite e ligados durante o dia).

Se o fim que se quer atingir é detecção de cheias em zonas habitacionais, estamos na presença de uma rede reactiva, pois o evento desplota a recolha de dados que, no entanto, tem que ser feita em tempo real, pois a tomada de decisões dependem desses dados. Quando é

necessário gerir a alimentação dos equipamentos que efectuem a recolha de dados torna-se importante controlar a energia consumida.

Esta imposição leva a que o envio dos dados recolhidos possa ser feita de várias maneiras. Se o equipamento que faz a recolha dos dados possui alimentação própria, pode envia-los continuamente. Caso contrário, se há necessidade de poupar energia os dados podem ser enviados de 5 em 5 horas, em situações em que ocorram eventos de interesse (por exemplo fogo) ou quando solicitados pelo utilizador.

O utilizador só pode solicitar o envio de dados se o sistema o permitir ou seja se o nó permita comunicação *half-duplex* ou *full-duplex*, havendo no entanto equipamentos cuja função seja unicamente transmissão de dados.

#### **2.1.4 Redes de Sensores Vs Redes Tradicionais de Dados**

Numa primeira abordagem, poder-se-ia pensar que as redes de sensores e as redes de dados ditas tradicionais (entre pc's por exemplo) seriam idênticas. No entanto, as especificidades de aplicações das redes de sensores, atribui-lhe necessidades que as diferenciam em dois vectores: energia e tipo de tráfego [23].

O consumo de energia assume-se como característica fulcral de diferenciação entre estes dois tipos de redes. Nos sensores *wireless*, a autonomia energética é determinante visto os dispositivos geralmente não estarem ligados a fontes de alimentação regulares (permanentemente alimentadas), necessitando assim de minimizar o consumo de energia.

Optimizações, na altura do desenho dos sensores, como a redução da capacidade de processamento, redução da memória e diminuição da largura de banda utilizada são compromissos comuns no mercado. Em sistemas mais recentes, a autonomia é aumentada recorrendo a painéis solares, mas não deixando de ser crítico o factor alimentação.

Numa rede tradicional de dados, os utilizadores estão ligados a um nó ou a um grupo de nós e requerem um serviço a um outro nó. A rede age como um meio de propagação da comunicação entre os dois. Um pedido de comunicação por parte do utilizador pode resultar numa resposta da outra parte.

Uma rede de sensores funciona de uma maneira um pouco diferente, no sentido que os nós são entidades independentes, mas fortemente ligadas, agindo como sistemas distribuídos. Os nós trabalham em conjunto de forma a produzir informação enriquecida. O utilizador raramente está interessado numa leitura de um determinado nó, mas sim em alguns parâmetros de um sistema físico, obtidos a partir do trabalho colaborativo de vários nós. Para que consiga esta informação, os nós formam um sistema aplicacional distribuído de forma a

responder ao utilizador. Os nós envolvidos no processo de envio de informação ao utilizador, informam sempre que detectem alterações ao fenómeno físico observado. Um pedido de comunicação por parte do utilizador poderá resultar numa resposta da aplicação distribuída/ serviço (o conjunto do trabalho de parte ou totalidade dos nós).

## **2.2 Redes de Sensores Ambientais**

As aplicações em sistemas ambientais passam por identificar mudanças de *habitat*, reportar eventos pouco comuns ou sazonais como por exemplo a migração das aves, identificar emergências como fogo e a prevenção da sua expansão e tarefas de *event notification* com o envio de alarmes em áreas de água que se encontrem contaminadas.

Auxiliar a comunidade científica a entender o comportamento da fauna e flora de determinada região, através da monitorização de *habitats* naturais, ecossistemas remotos em áreas de difícil acesso é um dos objectivos da utilização de uma RS nesta área.

Através da recolha de dados áudio, movimento, temperatura, vento, etc..., biólogos, ecologistas e cientistas em geral monitorizam o comportamento animal através da observação da vida selvagem (mapa de biodiversidade), identificam mudanças em determinados *habitats* e detectam micro-climas.

A protecção da natureza passa pela prevenção e detecção de fogos florestais que é assegurada por bombeiros e protecção civil. Dados sobre a localização precisa de um fogo e a direcção em que se propaga permite que a coordenação e orientação no terreno por parte destas forças seja mais eficaz bem como a determinação da sua causa (fogo posto, queimadas, relâmpagos). Esta tarefa pode ser executada por um conjunto de nós fixos (estações) e agentes móveis (UAV e outros sistemas móveis terrestres) que poderão dar indicações sobre a melhor forma de combater o incêndio conduzindo-os a fim de evitar zonas de fogo. Em alguns países a actividade sísmica pode ser detectada com maior precisão usando redes de sensores equipados com acelerómetros. [1]

A baixa produtividade dos terrenos condicionada pelo clima irregular, solos pouco profundos e a baixa fertilidade leva a que as RS sejam vistos como formas de rentabilizar as parcelas e diminuir os custos associados à produção agrícola. [1],[7],[12]

Assim o agricultor pode saber qual o impacto da variabilidade dos solos na produção de uma determinada cultura, saber quais as parcelas onde tem maior rentabilidade bem como saber se a aplicação de fertilizantes, pesticidas e a irrigação são feitas da melhor maneira. [53]. Com estes dados o proprietário consegue obter mapas com as necessidades (fertilizantes, pesticidas e água) específicas de cada parcela para que a produtividade seja máxima.



A protecção contra doenças e pragas nas plantações é outro dos problemas que os agricultores gostariam de prever ou seguir a sua evolução através da monitorização de condições atmosféricas e ambientais (Humidade, Temperatura, Precipitação, Exposição solar e velocidade do vento) .

Diversos sistemas de cercas convencionais são actualmente utilizados para controlar a localização dos animais. Estes sistemas incluem barreiras (tais como cercas de arame farpado) que bloqueiam a circulação física do animal e cercas eléctricas que utilizam o choque eléctrico para controlar o animal. A utilização de cercas eléctricas tornou-se mais generalizada, devido ao seu baixo custo e facilidade de deslocalização no entanto o agricultor pode ter um sistema de controlo de animais sem vedação utilizando informações localização GPS. Os sinais GPS são processados para fornecer informação sobre localização, que é comparada com os parâmetros desejados fronteira. Se o animal foi transferido fora da área desejada, uma estimulação é activada.

Controlar os parâmetros hidrológicos, químicos e biológicos da água de forma a garantir a sua qualidade para consumo doméstico e a recolha de amostras, particularmente em locais remotos ou sob condições adversas é outra das áreas de aplicação. Como medida preventiva a utilização de uma rede de sensores pode ser empregue na detecção de contaminação de solos e aquíferos e a sua evolução espaço-temporal alertando as autoridades e mesmo sugerindo a captação noutros aquíferos.

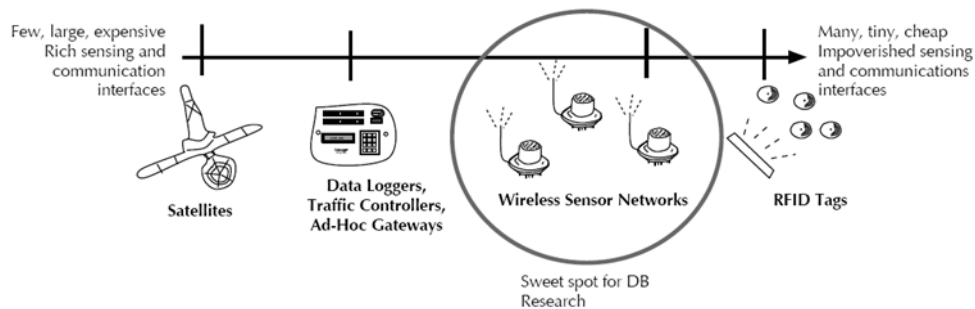
Com um conjunto de bóias (colocadas em estuários, rios e mar) podem ser detectados níveis perigosos de bactérias na água, dando facilmente o alerta a surfistas, banhistas e pescadores. [1]

As marés negras provêm de derrames de petróleo e consistem numa enorme mancha de crude que cobre uma determinada superfície aquática. A velocidade de propagação do crude depende de vários factores, como o estado do mar, as condições climáticas, a temperatura, o tipo de óleo.

A poluição atmosférica, nas regiões urbanas, tem vindo a aumentar devido sobretudo ao aumento do número de veículos em circulação, tornando-se por isso necessária informação sobre a qualidade do ar urbano. Além deste, as áreas urbanas tem problemas de ruído ambiental que poderiam ser colmatadas com a introdução de barreiras acústicas conhecendo-se para isso o mapa de ruído da cidade.

### 2.2.1 Exemplos de redes de sensores ambientais

Uma rede de sensores varia muito na sua dimensão e função. [25] Os dispositivos variam desde os embutidos em satélites (poucos, grandes e caros) a etiquetas de radiofrequência (RFID) (numerosas e baratas) como mostra a figura.



**Figura 9 - Espectro de dispositivos de sensores [25]**

Em [25] são classificadas as redes de sensores em quatro grupos diferentes segundo alguns critérios.

### 2.2.2 Redes de larga escala unifuncionais

Estas redes são normalmente usadas para cobrir grandes áreas geográficas, tendo apenas uma única finalidade e cujos nós sensores são caros e grandes. Nós que podem medir mais do que uma variável e em conjunto formam uma rede podendo ir desde uma simples estação meteorológica a um nível mais complexo como uma rede global sismográfica.

Normalmente têm alimentação própria, são cablados e o seu tamanho não é um factor importante.

### 2.2.3 Redes de sensores locais multifunção

Estas redes incluem tipicamente pequenos nós para monitorizar o ambiente. Estes dispositivos medem parâmetros relativamente simples (temperatura, humidade, etc...) e podem ser usadas numa variedade de aplicações, contudo, são mais susceptíveis de formarem sistemas *ad-hoc* e utilizam uma rede inteligente (nós com pouco poder computacional). Tipicamente, estes sistemas medem simples propriedades ambientais e meteorológicas.

### 2.2.4 Redes de biosensores

Conjunto de sistemas emergentes que se distinguem por usar biotecnologia. Estas redes distinguem-se principalmente pelo facto de tentar compreender um elemento biológico

usando sensores. Um biosensor inclui um elemento de sensorização biológica que é medido através de um sensor que pode ser electroquímico, óptico-electrónico ou acústico.

Genericamente estas redes de sensores ambientais requerem o uso de novas tecnologias para testes de ar, água ou solo. Actualmente, muitos poluentes são monitorizados *in situ*, mas são recolhidas e posteriormente analisadas em laboratório.

### **2.2.5 Redes de Sensores Heterogénea**

Estas redes de sensores integram os dados de diferentes tipos de dados descritos acima e monitorizam ambientes em diferentes escalas.

## **2.3 Redes de sensores ambientais - Cenários de aplicação**

### **2.3.1 Great Duck Island (GDI) System**

Nas ilhas remotas do *Great Duck*, cuja área é cerca de 1 Km<sup>2</sup>, aves como a *Leach's Storm Petrel*, embora vivam no mar, reproduzem-se em terra, construindo os seus ninhos em tocas subterrâneas e agrupadas em determinados pontos da ilha. No sentido de estudar o que leva a ave a escolher um determinado local para reproduzir a Universidade de *Berkley* instalou uma rede de sensores que fornece dados da temperatura, pressão atmosférica, humidade e infravermelho (para detecção de presença das aves nas tocas) [2].

A rede, numa primeira fase foi constituída por 32 nós sensores (*Mica Mote*) desenvolvidos pela própria universidade e que individualmente são caracterizados por uma baixa capacidade de processamento e energia.

Para além destes a rede possui algumas *motes*, colocadas na área de interesse científico e dentro de tubos, com um maior poder computacional pois neles são utilizados microfones e câmaras.

A entrada e saída de aves *Petrel* das tocas acontece durante a fase do amanhecer e anoitecer, logo não é necessário recolher amostras durante um ciclo de 24 horas. Assim sendo medições de 5-10 minutos durante estas duas fases são necessários. Esta taxa de amostragem é alterada durante o período de acasalamento, onde é medida as condições dentro e fora da toca sendo essa informação enviada a cada 2-4 horas.

As *motes* formam uma rede *ad hoc* permitindo a troca de informação entre si e dessiminam a informação através de *multihop*.

A *gateway*, cuja alimentação é assegurada por baterias e painéis solares, coordena a actividade dentro da rede, trata a informação proveniente dos sensores e armazena-a .

Cada rede criada tem uma *gateway* que possui uma antena externa direccionada que envia os dados para a estação base que se encontra aproximadamente a 100m.

A estação base por sua vez coordena a *gateway* e é quem fornece serviço à base de dados enviando dados a cada 15 minutos através de satélite.

O utilizador pode ter acesso à informação através do servidor de base de dados que se encontra em *Berkley*, enquanto que o responsável da rede pode alterar os parâmetros de forma a comunicar directamente com a mote, obtendo leituras actualizadas.

Posteriormente foram acrescentados mais nós sensores e 25 mini estações meteorológicas com uma autonomia expectável de 4 meses.

### **2.3.2 *PODS – A remote ecological micro-sensor network***

O *PODS* [48] é um projecto da Universidade do *Hawai* que combina a videovigilância de algumas espécies de plantas, raras ou em vias de extinção e a monitorização intensiva das condições atmosféricas nas áreas remotas onde essas plantas se encontram.

A necessidade de detectar em tempo real eventos importantes tais como a visita de polinizadores, consumo por parte de herbívoros e até mesmo visitas humanas bem como o acompanhamento das condições climáticas levou ao desenvolvimento de uma rede de sensores utilizando uma série de *PODS* (dispositivos sensores). Através das imagens conseguimos ver o comportamento das plantas (quando crescem, quando florescem, etc...).

Com uma densa rede de instrumentos meteorológicos, mede-se não só o que está a acontecer no local das plantas raras mas também nas zonas adjacentes.

Para o projecto desta rede havia a necessidade de usar *PODS* relativamente baratos, flexíveis e de fácil mobilização. Deveriam ser igualmente robustos pois terão de suportar uma grande variedade de condições ambientais extremas, com locais onde a precipitação é altamente ácida e outros onde facilmente temos temperaturas negativas.

Os nós sensores (*PODS*), são camuflados e encontram-se depositados no Parque Nacional de Vulcões do *Hawai*. Estes dispositivos possuem fonte de energia, computador, GPS, rádio *transceiver* e uma ficha permite a ligação de sensores meteorológicos.

Assim sendo nesta rede existem 3 tipos de *PODS*: *POD* de comunicação, *POD* especializado e Estação Base. O tipo mais vulgar é o *POD* de comunicação pois possui todas as funções de comunicação da rede e permite a ligação de instrumentação meteorológica. Os

*POD* especializados para além das funções de comunicação diferem fisicamente pois possuem câmaras digitais de alta resolução. Estes dispositivos requerem muito maior taxa de transmissão de dados e acarretam um maior consumo de energia. A estação base possui funções de comunicação como os outros *POD* no entanto serve como local de armazenamento temporário e posterior processamento de dados. Como serve para ligação à Internet este dispositivo não é usado para recolha de dados, ficando por isso longe da área a ser estudada.

Cada *POD* pode também ser utilizado para o envio de dados provenientes de outros *PODS*. O alcance de cada nó é relativamente curto, logo os *PODS* são colocados de forma a que os seus alcances se sobreponham, para que quando um nó falhar este seja substituído.

Temos portanto dois tipos de dados provenientes da rede: relativos a condições meteorológicas, cuja recolha é feita de dez em dez minutos e imagens que são recolhidos uma vez por hora.

### 2.3.3 Projecto *CORIE*

O projecto *CORIE* [46] desenvolvido pelo *Oregon Graduate Institute*, é um sistema de observação ambiental utilizado no estuário do rio Colombia para servir fins científicos e industriais e integra uma rede de sensores em tempo real, um sistema de gestão de dados e um conjunto de avançados modelos numéricos.

Esta rede é constituída por treze estações fixas, localizadas ao longo do estuário do rio, e uma embarcação (nó móvel) que garante mobilidade ao projecto. As estações fixas podem ser bóias ou sistemas fixos de amarras.

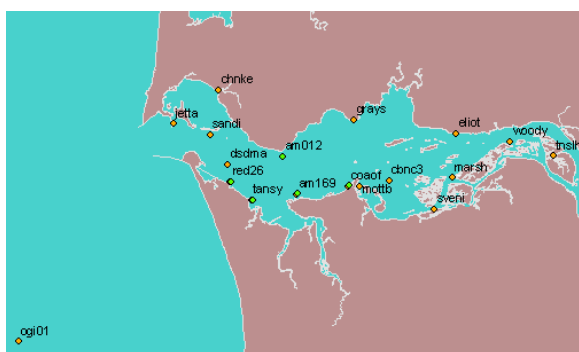


Figura 10- Localização espacial das estações



Figura 11- Sistema móvel marítimo

A maioria das estações possui CTD(*Conductivity, Temperature and depth*) para a medição da salinidade, temperatura e nível da água do rio, enquanto que a determinação da velocidade da água é assegurada por um conjunto de estações equipadas com perfiladores acústicos (efeito *doppler*). Parâmetros atmosféricos como velocidade e direcção do vento, temperatura do ar, humidade relativa e radiação são também aferidos.

Os dados recolhidos nas estações são transmitidos através de ligação *Wireless* a uma estação base que os transmite para um servidor centralizado e onde servem para alimentar complexo modelo ambiental. [3]

Este modelo é utilizado para assegurar melhorias na navegabilidade do rio, gestão de energia proveniente de hídricas bem como no estudo (passagem do salmão) e restituição de habitats.

A alimentação e fixação de antenas nos nós sensores (bóias) e a dificuldade de transmissão de sinal por estas, pois o sinal é frequentemente interrompido devido ao facto de, por vezes, a altura das ondas ultrapassar a altura da antena, são duas das dificuldades operacionais do sistema.

Os dados são amostrados em intervalos que variam de 1 a 15 minutos. A maioria das estações pode ser acedida em tempo real a dados recentes. A telemetria, em tempo real também pode ser feita via rádio.

Como ferramenta científica, *CORIE* foi projectada para compreensão de estuários e plumas em rios, pois permite caracterizar e prever a complexa circulação e processos de mistura num sistema que abrange a zona da foz e do estuário do rio Colombia.

### **2.3.4 Projecto ORCA**

Monitorizar a qualidade da água, a sul de *Puget Sound*, é o objectivo do projecto *ORCA* desenvolvido pela Universidade de *Washington* que utilizando uma variedade de sensores químicos e ópticos, permite a observação de uma área de potencial risco para o impacto da *eutrophication* (crescimento excessivo de uma planta aquática) indicando assim a presença no ecossistema aquático de nutrientes como azoto e fósforo.

Este projecto é constituído por quatro bóias autónomas, cada uma delas constituída por três partes principais: Amarras, plataforma e uma *box* subaquática.

A plataforma que se encontra à superfície é constituída por uma estação meteorológica, telefone celular, baterias, guincho eléctrico e painéis solares, sendo tudo controlado por um microcomputador com *S.O. Linux*. A caixa subaquática possui um CTD, eléctrodos para medição de oxigénio dissolvido e medição de flúor.

Em intervalos regulares de amostragem (cada 2 a 6 horas) [41], o microprocessador faz descer um guincho eléctrico e através da pressão proveniente do CTD, faz a monitorização de diferentes parâmetros numa coluna de água.

Os dados são gravados no computador e enviados de volta para o laboratório quase em tempo real através de comunicações móveis.

### ***2.3.5 ALERT - Automated Local Evaluation in Real Time***

No sentido de prever e prevenir inundações em áreas habitacionais, através da construção de uma carta de inundáveis foi desenvolvido, na década de 70 pelo Serviço Nacional de Meteorologia Norte Americano, o projecto *ALERT* [41] que através de uma previsão confiante de inundação através de uma detecção antecipada, permite reduzir assim o número de feridos, mortos e danos materiais provocados por este tipo de catástrofe natural.

A rede *ALERT* é constituída por estações (meteorológicas e de medição de caudal), repetidores e estação base. O *ALERT* permite monitorizar em tempo real, durante tempestades, os níveis de pluviosidade da região através de estações meteorológicas que para além desses dados dá a indicação de outras grandezas como temperatura, humidade etc..., enquanto que o nível do caudal está a cargo de estações colocadas no leito dos rios. Estas estações de recolha de dados transmitem a sua informação através de rádio o que leva a que, para além destas, existam na arquitectura do *ALERT* repetidores rádio e uma estação base que recebe todos os dados transmitidos pelas estações [3].

Cada estação (DCP) é programada para transmitir informação sempre que se desencadear alterações ambientais (por exemplo 1mm de precipitação e alteração de profundidade).

Um evento típico do sistema *ALERT* é a indicação da acumulação de 1mm de chuva no balde do pluviómetro. A estação envia uma mensagem de 4 *byte* usando FSK (Modulação por *shift keying*) a 300 *baud*. As plataformas de recolha de dados enviam, durante 133 milisegundos, 13 *bits* (0 a 8191) que identificam o sensor e 11 *bits* (0 a 2047) que codificam os valores recolhidos em dados.

No local de recepção dos dados o *software* identifica o local através do *ID* transmissor e decodifica os dados indicando um valor nas unidades apropriadas usando informação armazenada sobre o sensor.

A previsão temporal e espacial de uma inundação, bem como a sua magnitude, é feita recorrendo ao modelo *SYR* (*Santa Ynez River Flood Control Model*) permitindo a

coordenação com os serviços de emergência, serviços de protecção civil e meios de comunicação social.

Para além da detecção atempada de inundação este tipo de sistemas permite fazer a gestão e planeamento de recursos hídricos [3] .

### **2.3.6 *James Reserve Extensible Sensing System***

Nas Montanhas da *James Reserve* foi desenvolvido pela Universidade da Califórnia, uma rede de monitorização ambiental para detecção de microclimas no solo e sub-solo e detecção animal em mais de 100 locais com 12 hectares de área de estudo.

A rede de sensores na *James Reserve*, são o exemplo de uma rede hierárquica, possuindo nós que recolhem e executam um pequeno processamento (micro-nós) sobre os dados, enquanto que os *cluster-head* recebem esses dados e fazem-lhe um tratamento adicional. Estes macro sensores, para além de funcionarem como *cluster head*, fazem funções de gateway e no caso de um falhar o *routing* dos dados é feito pelo *cluster head* activo mais próximo. [6]

Nós sensores individuais, cada um podendo ter 8 sensores, são dispostos ao longo de uma linha que achesse os principais ecossistemas da reserva. Essa linha é constituída por 3 malhas, contendo cada malha 25 a 100 nós, que transmitem os dados através de *wireless* e são alimentadas através de painéis solares.

Nós móveis também são integrados na rede para visualização remota, através de imagens de alta resolução, e para recolha de dados em lugares onde existem “lacunas” dentro da rede de monitorização.

Dados e tempos de comunicação são normalizados e enviados para a Internet num subsistema *middleware* do tipo publicador-subscritor chamado *Subject Server Bus* (SSB) através do qual os dados são transmitidos para um tipo de clientes heterogéneos para o processamento e análise do histórico e dos dados em tempo real.

### **2.3.7 *FloodNET***

O objectivo principal da *FloodNET* é dar um aviso, pelo menos duas horas antes de acontecer uma inundação, de modo a que as acções possam ser tomadas para atenuar os riscos para as pessoas e bens. O *FloodNET* monitoriza o nível de água do rio *Crouch* em Inglaterra utilizando vários nós sensores *wireless* que formam uma rede de sensores.

Actualmente, é composto por doze nós, cada um dos quais inclui uma *board* de PC, com *wireless LAN PC card* para receber ou transmitir dados, um sensor de profundidade e



dois conversores analógico-digitais. Os dados são recolhidos e armazenados em memória com um consumo muito baixo.

Devido a obstruções como ondas, árvores e edifícios no campo do *FloodNET* não existe ligação directa entre os nós e a estação base, no entanto estes nós são capazes de comunicar a distâncias entre os 600 e os 800 metros, logo um nó envia os dados ao seu vizinho que por sua vez o transmite à estação base.

Os nós sensores recolhem informação das variáveis ambientais locais e dão o *feedback* para a Universidade de *Southampton* de modo a que possa ser usado um modelo para simulações da inundação para posteriormente fornecer um aviso adiantado de ameaça de inundação.

Esta página foi intencionalmente deixada em branco.

## 3 Estimação espacial

### 3.1 Introdução

Compreender a distribuição espacial de dados oriundos de fenómenos ocorridos no espaço constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento, seja em saúde, ambiente, geologia, agronomia, entre outras. [19]

Quando não se possui um modelo que descreva um determinado fenómeno físico não conseguimos saber qual ou quais os valores fora das posições espaciais amostradas. Os problemas de natureza ambiental são demasiado complexos para serem explicados através de modelos deterministas à excepção existe na evolução de plumas de contaminação. Pelo facto de não se poder repetir a amostra ou observação num espaço e no tempo, os fenómenos relacionados com as ciências da terra e do ambiente, possuem um grau de incerteza elevado ligado ao seu conhecimento. [18]

Todas estas aplicações fazem parte da análise espacial de dados geográficos. A ênfase da análise espacial é mensurar propriedades e encontrar relações entre variáveis, levando em conta a localização espacial do fenómeno em estudo de uma forma explícita. Ou seja a ideia é incorporar ao espaço a análise que se deseja fazer.

Uma das dificuldades apresentadas para a monitorização ambiental é a recolha ou leitura de dados de sensores instalados em campo. A recolha manual dos dados históricos de um sensor usando um *datalogger* ou *laptop* não permite uma leitura instantânea ou em tempo real simultaneamente em diversos pontos. A interligação entre sensores e uma central usando cabos apresenta um custo elevado em função de grandes extensões e da estrutura de instalação destes cabos (*dutos, postes*), além de dificultar o deslocamento das unidades sensoras.

A utilização de um meio de comunicação sem fio contribuiu em muito para as necessidades da monitorização ambiental, possibilitando o emprego de vários conjuntos de nós sensores de baixo custo.

No entanto o emprego de apenas um sensor, não fornece maiores informações sobre a dinâmica associada ao evento, apenas a informação pontual específica. Essa situação pode ser atenuada deslocando-se o sensor (adicionando um custo elevado), mas mesmo assim a informação obtida é semi-dinâmica considerando-se os tempos de deslocamento envolvidos.

Por outro lado uma RS possibilita criar um novo método de monitorização “virtual”, onde mesmo em que um determinado ponto não possua um sensor, a proximidade de outros sensores e análise dos dados permitem estimar a sua leitura. [30]

A precisão e exactidão das medidas dependem sempre da capacidade própria de processamento do nó, pois o processamento local (cálculo, conversão, estimação, comparação de dados, etc) evita sobrecarga de dados na rede permitindo que o envio de dados para a central seja feito quando ocorre um evento (*event notification*) ou quando solicitado.

A distribuição dos nós sensores não deverá ser feita de uma forma aleatória e irregular pois corre-se o risco de vários nós acabarem por medir o mesmo dado, tornando a rede altamente redundante.

### **3.2 Geoestatística e suas aplicações**

Inicialmente, métodos de estimação espacial foram desenvolvidos para solucionar problemas de mapeamento geológico, mas o seu uso expandiu-se com sucesso no mapeamento de solos, mapeamento hidrológico, mapeamento atmosférico e outros campos relacionados [31].

A ideia antiga de abundância dos recursos hídricos passada entre as gerações está sendo reestruturada a fim de inibir o desperdício e a destruição dos mananciais. [35]

As barragens geram uma redução das velocidades da corrente. Provocam a deposição gradual dos sedimentos transportados pelos cursos de água e consequente assoreamento, diminuindo a capacidade de armazenamento dos reservatórios e podendo inviabilizar a operação de aproveitamentos hidroeléctricos, além de ocasionar problemas ambientais de diversas naturezas.

O conhecimento sobre a espacialidade das chuvas possui relevância para estudos tais como a previsão da disponibilidade hídrica no solo, cálculos do balanço hídrico ou a modelação distribuída de vazão que pode subsidiar projectos de construção de barragens de hidroeléctricas ou controle de cheias e secas. [33]

O estudo da variabilidade espacial do potencial de erosão das chuvas [42], através de recolha de dados de pluviómetros, permite a estimação de perdas de solos servindo como guia na selecção de sistemas de cultivo e práticas de conservação do solo em geral.

Os atributos físico-hídricos dos solos são de reconhecida importância no desenvolvimento das culturas agrícolas, sendo desejável sua caracterização, modelagem e mapeamento. Nas últimas décadas, conceitos de monitorização e gestão do processo de produção agrícola vêm sendo amplamente discutidos e aplicados, gerando grande quantidade de informação, obtida pelo estudo dos factores que interferem na produtividade agrícola. Tais conceitos levam em consideração a variabilidade espacial de variáveis georreferenciadas,

estudada pela geoestatística, principalmente aquelas relacionadas aos atributos físicos e químicos do solo. [34]

Inferir um índice de produtividade de uma espécie de árvore para os locais onde esta espécie ocorre [32], mas onde não se procedeu à recolha de informação de campo é uma ferramenta importante para o planeamento e gestão dos recursos florestais de uma região ou país.

Perdas no transporte de produtos alimentares, devido à má gestão da temperatura e da deterioração da qualidade pode atingir até 35% [31]. Uma pesquisa anterior, com um elevado número de pontos de medição de temperatura espacial revelou desvios de mais de 5 ° C no transporte refrigerado dentro de um camião.

Essas perdas podem ser atenuadas através de uma melhor fiscalização da cadeia de frio recorrendo a redes de sensores sem fio mas o número de sensores tem que ser reduzida a um valor economicamente útil para uso no transporte regular.

### **3.3 Representação geométricas de dados**

Na geoinformação, computadores são utilizados como instrumentos de representação de dados espacialmente georreferenciados. Neste capítulo analisa-se os vários tipos de representação de dados geográficos.

#### **3.3.1 Tipo e natureza de dados em análise espacial**

Os problemas de análise espacial consideram três tipos de dados: eventos ou padrões pontuais, superfícies contínuas e áreas com contagens e taxas agregadas. [19]

Nos eventos ou padrões pontuais, os fenómenos são expressos através de ocorrências identificadas como pontos localizados no espaço como por exemplo localização de crimes. As superfícies contínuas são estimadas a partir de um conjunto de amostras de campo, que podem estar regularmente ou irregularmente distribuídas. Normalmente, este tipo de dados é resultado de levantamento de recursos naturais que incluem mapas geológicos, topográficos, ecológicos e fitogeográficos. As áreas com contagens e taxas agregadas são áreas normalmente delimitadas por polígonos fechados onde se supõe haver homogeneidade interna ou seja, mudanças importantes só ocorrem nos limites.

### **3.3.2 Representação computacional dos dados geográficos**

O termo Sistemas de Informação Geográfico (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos e armazenam a geometria e os atributos dos dados que estão georeferenciados isto é, localizados na superfície terrestre e representados numa posição cartográfica.

As representações geométricas mais utilizadas em SIG são pontos, polígonos, amostras, malha regular e imagem.

Um ponto 2D é um par ordenado  $(x,y)$  de coordenadas espaciais e dá-nos a indicação do local de ocorrência de um evento. Por sua vez um polígono é um conjunto de pares ordenados  $(x,y)$  de coordenadas espaciais, de tal forma que o último ponto seja idêntico ao primeiro, formando uma região fechada de plano. As amostras consistem em pares ordenados  $(x,y,z)$  nos quais os pares  $(x,y)$  indicam as coordenadas geográficas e  $z$  indica o valor do fenómeno estudado para essa localização.

A malha regular não é mais do que uma matriz onde cada elemento está associado a uma valor numérico. Esta matriz está associada a uma região da superfície terrestre, a partir da coordenada inicial, normalmente referenciada ao canto inferior esquerdo da matriz, e de espaçamentos regulares nas direcções horizontal e vertical. A imagem, é representada igualmente como uma matriz onde cada elemento está associado a um valor inteiro (normalmente na faixa entre 0 e 255), utilizada para visualização.

### **3.3.3 Tipo de variáveis ou Natureza de dados**

Podemos classificar os dados quanto à sua natureza como contínuos ou discretos. Variáveis contínuas são aquelas que podem tomar um número infinito de valores reais tais como temperatura, cota de terreno, teor químico etc.... enquanto que as outras são podem tomar um número discreto de valores numéricos como por exemplo cor do solo, tipo de rocha, textura, etc....

Normalmente a referência geográfica dos dados está guardada nas coordenadas das estruturas de dados, que está associada a uma projecção cartográfica planar, ou a valores de latitude (coordenada  $y$ ) e longitude (coordenada  $x$ ).

O processo de estimação produz uma malha regular que descreve de forma aproximada o fenómeno na região em estudo. Esta malha pode ser transformada numa imagem para fins de representação.

### **3.3.4 Análise Estatística**

Antes de se proceder à avaliação e estimação do valor de uma grandeza, é recomendado que se faça uma análise estatística dos dados disponíveis ou recolhidos a partir de nós sensores. Essa análise é feita com o objectivo de caracterizar e descrever estatisticamente as distribuições dos dados para melhor entendimento do comportamento das variáveis de interesse. Um outro lado de suma importância no processo de avaliação é aquele que mede a dispersão dos valores numa determinada área. A dispersão desses valores está relacionada com erros de amostragem, erros na determinação desses valores e, sobretudo, com a variabilidade natural da grandeza.

#### **3.3.4.1 Estatística Descritiva**

A Estatística Descritiva, também chamada por vezes Análise Exploratória de Dados, desenvolveu um conjunto de medidas de síntese e de representações gráficas cujo objectivo é descrever e resumir um conjunto de dados e pesquisar as estruturas e anomalias presentes.

Regra geral os valores de um conjunto de dados estão dispersos; então é aconselhável ver as suas tendências, isto é, verificar se os valores têm tendência para se agruparem em torno de um valor “centro”. Essas medidas são designadas de medidas de localização ou de tendência central, pois representam os fenómenos pelos seus valores centrais, sendo a média a medida mais usada na vida prática.

Quando os dados de um conjunto não se encontram todos igualmente distribuídos em torno da média é necessário utilizar outras medidas estatísticas que nos dêem uma informação mais completa acerca da “forma” da distribuição. A amplitude, variância e desvio padrão são alguns exemplos desse tipo de medidas normalmente conhecidas por medidas de dispersão.

Até agora foram referidas formas de analisar informações referentes a uma única variável, porém frequentemente estamos interessados em analisar o comportamento conjunto de duas ou mais variáveis pois, além de medidas individuais, também pode ser de interesse conhecer se elas têm algum tipo de associação entre si. Uma forma bastante útil de se observar a relação entre duas variáveis é o gráfico de dispersão. O gráfico de dispersão dá-nos uma ideia do tipo de relacionamento entre duas variáveis.

Se a relação entre essas variáveis for do tipo linear, podemos medir esse grau de associação através de duas medidas - covariância e coeficiente de correlação- comumente designadas por medidas de associação. A covariância é uma medida do grau de correlação linear entre duas variáveis (a variação conjunta entre elas) e poderia ser utilizada para medir o

grau e o sinal de relacionamento entre duas variáveis, mas ela é difícil de interpretar pois varia entre  $+\infty$  e  $-\infty$ .

O coeficiente de correlação linear é igualmente uma medida da correlação linear entre duas variáveis sendo preferível à covariância por ser mais preciso e independente das unidades de medida de X e Y, variando entre -1 e 1. A covariância e o coeficiente de correlação medem a intensidade da relação linear (eventualmente) existente entre duas variáveis aleatórias.

#### **3.3.4.2 Variável aleatória**

Com o estudo das probabilidades pretende-se “medir a incerteza” de acontecimentos cujo resultado, à partida, não se consegue prever. Prever e calcular resultados do acaso e medir o grau de incerteza ou hipóteses é objectivo do cálculo de probabilidades.

Uma variável aleatória é uma grandeza que pode tomar uma série de valores possíveis, a cada um dos quais está associada uma probabilidade de ocorrência podendo ser definida como uma função X que associa a cada elemento s de S um número real X(s);

As variáveis aleatórias X são completamente caracterizadas pela sua função de distribuição de probabilidade  $F(x)$

$$F(x) = P(X \leq x) \quad (3.1)$$

Por outro lado a função densidade de probabilidade é uma função  $f(x)$  que caracteriza também a distribuição da variável aleatória X. Obtém-se por derivação da função de repartição: é a tangente de  $F(x)$  em cada ponto x.

$$f(x) = \lim_{dx \rightarrow 0} \frac{F(x+dx) - F(x)}{dx} = F'(x) \quad (3.2)$$

Para caracterizar e sintetizar a distribuição das variáveis aleatórias e a forma da densidade de probabilidade em alguns valores numéricos é comum usar os momentos de ordem k de uma variável aleatória X.



$$M^k(X) = \int_{-\infty}^{+\infty} f(x) x^k dx$$

*Variáveis contínuas*

$$M^k(X) = \sum_{\text{todas } x} x_i \cdot P(x_i)$$

*Variáveis discretas*

Os momentos de uma distribuição servem para caracterizar uma distribuição, não apenas no que se refere à sua centralidade e dispersão, mas também com relação a outras características, como simetria ou assimetria da densidade de probabilidade.

### 3.3.4.3 Média

O momento de 1ª ordem de uma variável aleatória é designado por esperança matemática ou valor esperado da variável X e é calculado como:

$$E(X) = \sum_{i=1}^n x_i \cdot P(x_i) \quad (3.3)$$

onde  $P(x_i)$  é a probabilidade associada à ocorrência da i-ésima variável X. Contudo, essa expressão é encontrada, frequentemente, na sua forma simplificada, na qual se assume que as probabilidades associadas às n variáveis aleatórias são iguais entre si e, portanto, iguais a 1/n [38]

$$E(x) = \bar{x} = \sum_{i=1}^n x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.4)$$

A esperança matemática corresponde à média quando todos os valores da variável são equiprováveis, sendo o seu significado físico o do centro de gravidade de um conjunto de pontos de massa  $f(x_i)$  com coordenadas  $x_i$ ; localiza pois a posição central da distribuição (é uma medida de localização central).

Uma das propriedades mais importantes associadas à média: é a que diz que a média de uma variável aleatória somada ou subtraída de uma constante é igual à média dessa variável somada ou subtraída da mesma constante.

$$E(x \pm k) = E(x) \mp k \quad (3.5)$$

### 3.3.4.4 Variância

Pode-se definir ainda outro tipo de momentos: os momentos centrados de ordem  $k$  de uma variável aleatória  $X$ .

$$M_c^k(X) = \int_{-\infty}^{+\infty} f(x) (x - \bar{x})^k dx$$

*Variáveis contínuas*

$$M_c^k(X) = \sum_{\text{todas } x} f(x) (x - \bar{x})^k$$

*Variáveis discretas*

O 2º momento centrado mede uma característica importante da sua distribuição - a dispersão em torno da esperança. Tem o significado físico de um momento de inércia e designa-se por variância é determinado por

$$\text{Var}(X) = \sum_{i=1}^n (x_i - \bar{x})^2 \cdot P(x_i) \quad (3.6)$$

Novamente, assumindo a probabilidade de ocorrência das  $n$  variáveis aleatórias iguais entre si, ou seja, iguais a  $1/n$ , tem-se que

$$\text{Var}(X) = S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \cdot P(x_i) \cdot \frac{1}{n} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \cdot P(x_i) \quad (3.7)$$

O desvio padrão é simplesmente a raiz quadrada da variância e é expresso na mesma unidade dos valores originais [20].

## 3.4 Teoria das variáveis regionalizadas

### 3.4.1 Análise Geoestatística

É impossível medir fenómenos como a quantidade de chuva ou a concentração de uma poluente em todos os pontos de uma área geográfica. No entanto se recolher uma amostra da grandeza em várias posições dentro da área de estudo, posso então usar estas amostras para fazer inferências em toda a área geográfica.

Um facto muito importante que deve ser levado em consideração, é a variação espacial que o dado sofre. Durante muito tempo estudou-se essa variabilidade através de cálculos utilizando parâmetros como média, desvio padrão, baseados na estatística clássica, considerando que esse tipo de variação era aleatório.

A estatística é o ramo da matemática que se dedica ao estudo da colheita, organização, representação e análise de dados e às medidas que permitem extrair conclusões e fazer inferências a partir desses dados.

Enquanto que a estatística descritiva dedica-se à representação e análise de um conjunto de dados, relacionados com uma amostra ou com uma população conhecida de forma exaustiva, a estatística inferencial dedica-se à previsão/estimação de parâmetros de uma população a partir dos dados de uma amostra colhida de forma adequada.

Os procedimentos gerais utilizados na estatística clássica passam por seleccionar um local típico ou representativo do fenómeno, retirando as amostras ao acaso para evitar tendências e evitar locais espacialmente variáveis para evitar erros. Para além destes as medidas são feitas em locais típicos que representem, por exemplo um solo particular o que leva a que a variabilidade seja má considerando-o como um erro. As hipóteses em que a estatística básica se baseiam são a de assumir uma distribuição de frequência e considerar que os erros de observações têm média zero.

A partir daí surge o conceito da Geoestatística, que leva em consideração a localização geográfica e a dependência espacial. De entre os objectivos desta área destaca-se a descrição do comportamento espacial dos dados, estimação de um valor desconhecido num determinado ponto do espaço, determinação do grau de incerteza associado às estimações e a utilização dos valores conhecidos de uma variável na estimação de outra variável.

A parte central da geoestatística é a ideia de que medidas mais próximas tendem a serem mais parecidas do que valores observados em locais distantes e fornece métodos para quantificar esta correlação espacial e incorporá-la na estimação e na inferência.

Em conclusão enquanto que na estatística clássica existem independências entre as amostras pois ignoram a posição geográfica da amostra, a geoestatística considera a dependência entre as amostras pois os locais de amostragem são importantes.

### **3.4.2 Conceitos Básicos em estatística espacial**

A maioria das propriedades das Ciências da Terra, variam continuamente no espaço. Como consequência, os valores em locais mais próximos no solo são mais similares dos pontos que estão mais distantes. Eles dependem uns dos outros no sentido estatístico.

O que quer nos dizer este princípio? Que nada na natureza (como na vida real) acontece por acaso. Se encontramos poluição numa parte de um lago, é provável que locais próximos a esta amostra também estejam poluídos. [36]

Um conceito chave na Análise Espacial é a dependência espacial, e sua formulação matemática, a autocorrelação espacial. Estas noções partem do que *Waldo Tobler* chama de "a primeira lei da geografia": "todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes"

Do ponto de vista matemático, o valor de uma propriedade do solo em qualquer lugar da superfície terrestre depende da sua posição. Deve existir alguma expressão matemática que descreve esta variação de um ponto para outro. Na prática esta variação é muito irregular, qualquer equação adequada relacionando valores de uma propriedade do solo com posição deve ser muito complexa

### **3.4.3 Dependência espacial e Autocorrelação espacial**

De uma forma geral, estes dados estão disponíveis na forma de amostras pontuais, e para utilizá-los de uma forma efectiva necessita-se de um procedimento de interpolação, para gerar uma representação na forma de malha regular. As amostras são valores representativos do fenómeno estudado, normalmente obtidas a partir do levantamento de campo, e que apresentam consistência de metodologia e unidade. Para gerar superfícies que aproximem o fenómeno estudado de forma realista, é necessário modelizar a sua variabilidade espacial.

Segundo *Noel Cressie* “ a dependência espacial está presente em todas as direcções e fica mais fraca à medida que aumenta a dispersão na localização de dados. A maior parte das ocorrências sejam estas naturais ou sociais, apresentam entre si uma relação que depende da distância. Se encontrarmos poluição num trecho de um lago é provável que locais mais próximos destes estejam poluídos. A presença de uma árvore inibe o desenvolvimento das outras, essa inibição diminui com a distância, e após um determinado raio outras árvores grandes serão encontradas. Computacionalmente o conceito de dependência espacial é obtido através da autocorrelação espacial. A correlação é um conceito estatístico para mensurar a relação entre duas variáveis aleatórias. “Auto” indica que a medida de correlação é realizada com a mesma variável aleatória, medida em locais distintos do espaço.

No entanto como avalio essa dependência? Utilizando indicadores todos com a mesma ideia: verificar como varia a dependência espacial, a partir da comparação entre os valores de

uma amostra e os seus vizinhos. Valores significativos de índices de auto correlação espacial são evidências de dependência espacial e indicam que o postulado de independência de amostras, base da maior parte dos procedimentos de inferência estatística, é inválido e que os modelos inferenciais para estes casos devem levar explicitamente o espaço em conta nas suas formulações

### **3.4.4 Características das variáveis regionalizadas**

Segundo *Bubeiennn& Haas* (1969), existem características qualitativas de variáveis regionalizadas que os métodos estatísticos não convencionais não conseguem reconhecer.

Uma variável regionalizada é numericamente definida por um valor, o qual está associado a uma amostra de tamanho, forma e orientação específicos. Essas características geométricas da amostra são denominadas suporte geométrico. O suporte geométrico poderá não necessariamente compreender volumes mas também a áreas e linhas. Quando o suporte geométrico tende a zero, tem-se um ponto ou amostra pontual e o suporte geométrico é imaterial.

Por exemplo para estudar a variação da saturação de água no solo, são recolhidas amostras de  $10\text{cm}^3$ . A variável regionalizada é a humidade do solo e o suporte geométrico é o volume da amostra ( $10\text{cm}^3$ ). Note que, nesta recolha, o conteúdo de água no solo depende não só da localização da amostra mas também de seu tamanho, forma e orientação. Uma amostra de forma cilíndrica e longa, retirada na vertical, contém mais água que uma amostra de mesmo tamanho e a forma retirada na direcção horizontal em relação à superfície do solo. Se o volume da amostra for  $10\text{ m}^3$  ao contrário de  $10\text{cm}^3$ , o resultado também será diferente.

Resumindo, a teoria das variáveis regionalizadas considera a geometria das amostras, distintamente da estatística clássica onde a forma, o tamanho e a orientação não são considerados. Um ensaio estatístico clássico como o lançamento de moedas têm resultados que são independentes se a moeda é grande ou pequena, leve ou pesada, e de como é lançada.

Outra das características das variáveis regionalizadas. A regionalização pode apresentar anisotropias quando apresenta variações graduais numa direcção e rápida ou irregular em outra.

A variação espacial de uma variável regionalizada pode ser, dependendo do fenómeno, grande ou pequena, mas deve existir uma certa continuidade ponto a ponto. Apesar da complexidade das flutuações, uma continuidade média está geralmente presente.

As variáveis regionalizadas, que representam os valores de variáveis referenciadas geograficamente, foram introduzidas para descrever quantitativamente variações espaciais em

jazidas de minério. Para o desenvolver o conceito de variável regionalizada e a sua continuidade considere-se as seguintes séries de amostras de uma grandeza

**Tabela 1- Série de amostras de uma grandeza**

<b>Série A</b>	1	7	3	6	2	9	4	8	5
<b>Série B</b>	1	3	5	7	9	8	6	4	2

As características estatísticas destas duas séries de números, medidas por meio da média e da variância (Tabelas 1 e 2) são idênticas, pois apresentam os mesmos valores

**Tabela 2- Estatísticas medidas para as duas séries de números**

<b>Série</b>	<b>Média</b>	<b>Variância</b>
A	5	6,67
B	5	6,67

Assim as estatísticas obtidas por métodos clássicos não conseguem reconhecer a diferença existente entre as duas séries em estudo, pois consideram as amostras independentes entre si. Com a simples soma das diferenças tenderia a anular-se, optou-se pela soma do quadrado das diferenças, que, dividido pelo número de pares, dá sentido a uma medida de variância com significado espacial, pois é dependente da distância utilizada.

Assim calculando-se a variância espacial para um e dois intervalos de amostragem para as amostras da série A

$$A = [(1 - 7)^2 + (7 - 3)^2 + (3 - 6)^2 + (6 - 2)^2 + (2 - 9)^2 + (9 - 4)^2 + (4 - 8)^2 + (8 - 5)^2] / 8 = 22$$

$$A = [(1 - 3)^2 + (3 - 2)^2 + (2 - 4)^2 + (4 - 5)^2 + (7 - 6)^2 + (6 - 9)^2 + (9 - 8)^2] / 7 = 3$$

Calculando-se a variância espacial até quatro intervalos de amostragem, tem-se que os resultados mostrados na tabela 3.

**Tabela 3- Variâncias espaciais para as séries A e B, determinadas até quatro intervalos de amostragem**

Intervalo de amostragem	Varição espacial A	Varição espacial B
1	22,00	3,63
2	3,00	12,86
3	23,67	23,83
4	3,80	29,60

Observando-se o gráfico da figura 12 verifica-se que a série A é muito errática, enquanto a série B é mais uniforme. Na série B, as variações espaciais aumentam conforme o intervalo de amostragem, pois a correlação entre os valores diminui com a distância. Uma maneira prática de verificar se há correlação espacial nos dados, é calcular a variância espacial para um intervalo de amostragem e compará-la com a variância amostral; se a variância espacial for menor, então há correlação, caso contrário não há.

**Figura 12- Variação espacial em função de intervalos de amostragem para as séries A e B [20]**

Veja-se, por exemplo, que na série B a variância espacial para um intervalo de amostragem é igual a 3,63 para uma variância amostral de 6,67, portanto com uma boa correlação espacial, enquanto para a série A a mesma variância é igual a 22,00, ou seja, não apresenta correlação espacial. [38]

Este exemplo enfatiza a importância da medida da continuidade espacial da variável regionalizada. Assim, torna-se necessário considerar a posição espacial relativa de cada uma das observações nas duas amostras, para que as mesmas sejam diferenciadas. A continuidade espacial da variável regionalizada pode ser analisada a partir do variograma que irá ser estudado mais à frente.

### 3.4.5 Variáveis regionalizadas

Quando uma variável  $Z(x)$  depende da localização espacial determinada pelas coordenadas  $x$  (num espaço a 1,2 ou 3 dimensões) chama-se variável regionalizada [21].

Esta definição de variável regionalizada é apenas descritiva. No entanto, como se pretende extrair conclusões quantitativas acerca da variável torna-se necessário um modelo capaz de descrever rigorosamente o comportamento deste tipo de variável [21].

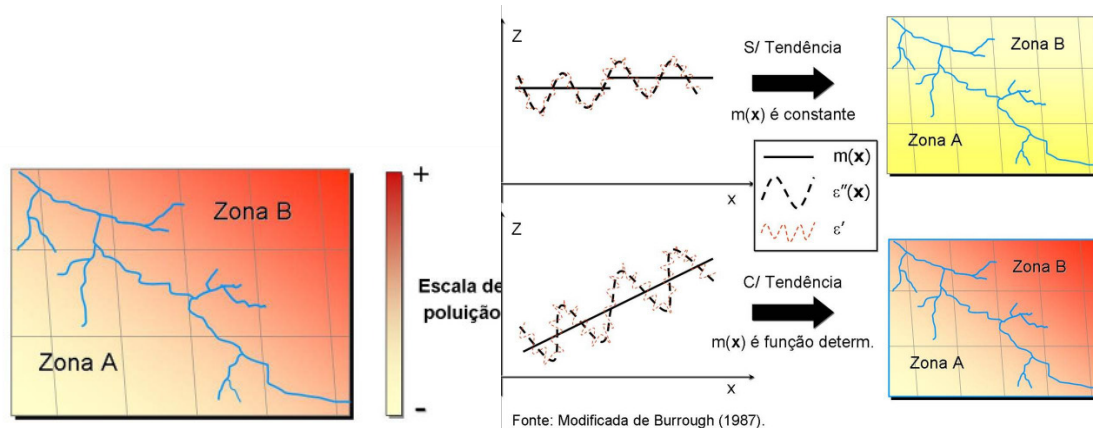


Figura 13- Componentes da variação espacial de uma variável regionalizada [45]

As variáveis naturais possuem propriedades intermediárias entre variáveis puramente aleatórias e as completamente determinísticas. Assim sendo numa área poluída, existem zonas que têm, em média, uma maior quantidade de metal pesado que em outras no entanto dentro de uma zona de uma área poluída, o conteúdo de metal pesado apresenta flutuações. Em suma estas variáveis apresentam um comportamento errático, a nível local e a escalas superiores um comportamento estruturado.

A variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes:

$$Z(x) = m(x) + \mathcal{E}'(x) + \mathcal{E}'' \quad (3.8)$$

- $x$  é o vector de coordenadas geográficas;
- $m(x)$  Expressa a tendência (“*Drift*”) do fenómeno. Pode estar associado a um valor médio constante (quando não há tendência) ou então ser uma função determinística;
- $\mathcal{E}'(x)$  é um termo estocástico, que varia localmente e depende espacialmente de  $m(x)$ ;



- $\mathcal{E}$  é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância  $\sigma^2$ .

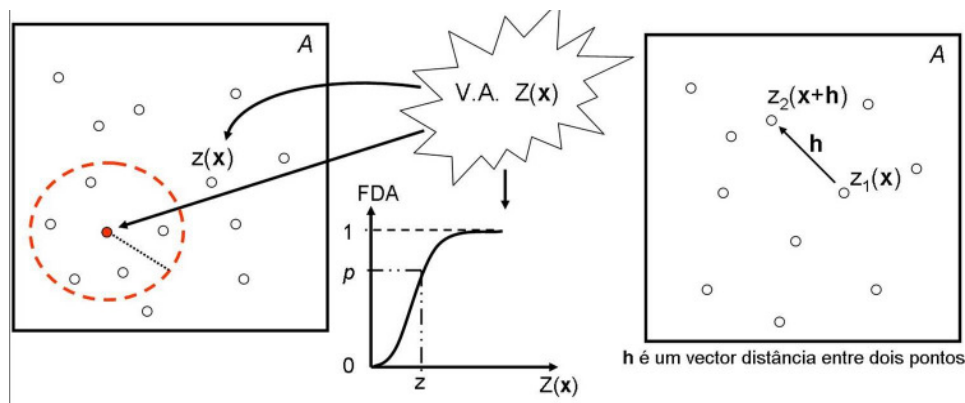
### 3.4.5.1 Funções aleatórias

O duplo aspecto, aleatório e estruturado, das V.R. pode ser interpretado com apoio na teoria das Funções Aleatórias (F.A.). Designa-se por  $Z$  a variável aleatória (que contém todos os valores possíveis) e por  $z$  uma realização particular dessa variável aleatória. [45].

Localmente, um valor  $z(x)$ , é interpretado como uma das possíveis realizações da Variável Aleatória (V.A.)  $Z(x)$ . O conjunto de V.As.,  $\{Z(x)\}$ , é uma Função Aleatória  $Z(x)$ .

Cada par de pontos em  $x$  e  $x + h$ , correspondentes às V.As.  $Z(x)$  e  $Z(x + h)$ , não são independentes, mas estão relacionados por uma correlação expressando a estrutura espacial do fenómeno. Essa correlação é a covariância

$$C[Z(x), Z(x+h)] = E[Z(x) \cdot Z(x+h)] - E[Z(x)] \cdot E[Z(x+h)] \quad (3.9)$$



**Figura 14 – Aspecto Aleatório de uma variável regionalizada [45]**

Este modelo incorpora os aspectos aleatórios e estruturado da variável regionalizada:

- Localmente, num ponto  $x$ ,  $Z(x)$  é uma variável aleatória;
- Dois pares de pontos  $x$  e  $x+h$ , distanciados de  $h$ , estão correlacionados, se houver estrutura espacial.

Uma função aleatória é caracterizada pela sua lei de distribuição, que pode ser definida pela distribuição conjunta das variáveis aleatórias que a definem. Ora a variável regionalizada, mesmo que fosse totalmente conhecida, é a única realização disponível, colocando-se então a questão de como inferir aquelas leis de distribuição ou mesmo de alguns

primeiros momentos. Para tornar o modelo útil (operacional), torna-se necessário impor algumas condições ligadas à homogeneidade espacial. Estas hipóteses adicionais e restritivas, recorrendo a um certo grau de estacionariedade, permitem reduzir os requisitos de que dependem a função aleatória, tornando exequível a sua utilização na resolução de problemas práticos. [21]

### 3.4.5.2 Hipóteses consideradas ou restritivas

O paradigma que se estabelece, para inferir as Funções de Distribuição Acumuladas (FDA) e interpolar valores em localizações não amostradas, é o de assumir a hipótese de estacionariedade [45]. Diz-se que um processo (ou uma variável) é estacionária se o desenvolvimento desse processo no tempo ou no espaço ocorrer de maneira mais ou menos homogênea, com oscilações aleatórias contínuas em torno de uma valor médio, em que nem a amplitude média, nem as oscilações mudam bruscamente no tempo e no espaço [44]. Uma F.A. é dita ser estacionária no campo A se a sua FDA multivariada é invariante a qualquer translação  $l$  nas  $k$  coordenadas do vector  $x$ .

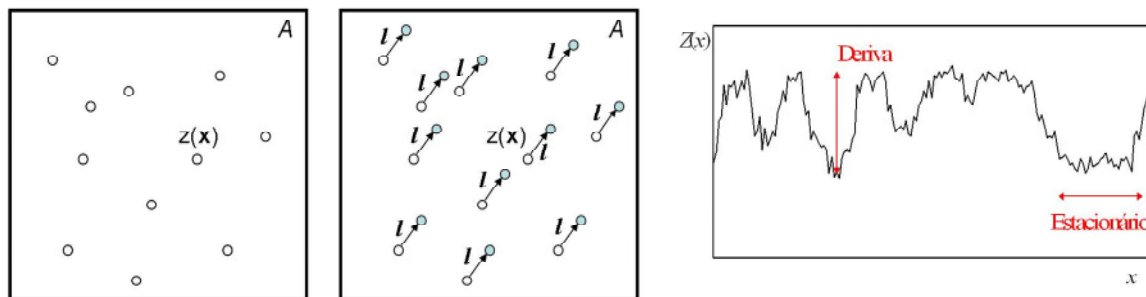


Figura 15 – Hipótese de estacionariedade [45]

Ou seja se a esperança matemática de uma variável aleatória é constante, independentemente da origem que se toma no espaço ou no tempo, podemos dizer que a variável é estacionária de primeira ordem e, portanto, a média será a mesma para todo o processo.

$$E[Z(x)] = m_1(x) = \text{constante} \quad (3.10)$$

Se um processo ou variável é não estacionário é porque não apresenta as características anteriormente citadas logo as características do processo dependem da origem que é tomada como referência.

A estacionaridade usada na Teoria das variáveis regionalizadas é a estacionaridade de segunda ordem que considera somente o primeiro e o segundo momentos invariantes da F.A..

Sob esta hipótese, admite-se que a componente determinística,  $m(x)$ , é constante (não há tendências na região). Então  $m(x)$  é igual ao valor esperado da variável  $Z$  na posição  $x$ , e a diferença média entre os valores observados em  $x$  e  $x+h$ , separados por um vector distância  $h$  é nula.

$$E[Z(x)] = m, \quad \forall x \in A \quad (3.11)$$

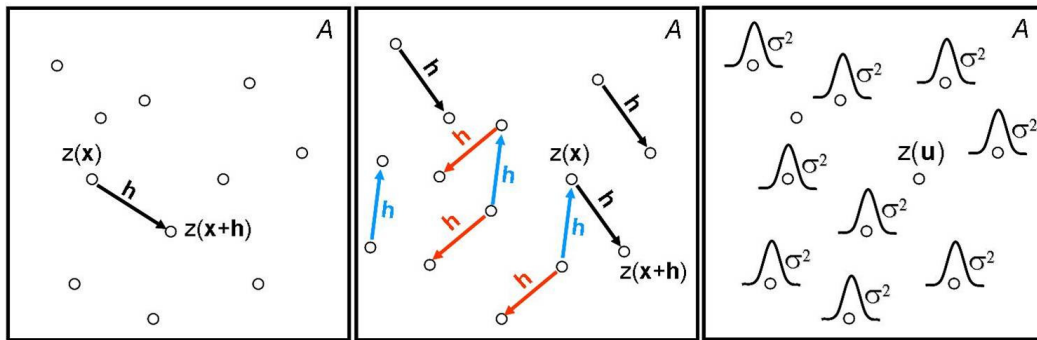
$$E[Z(x)] = E[Z(x+h)] = m \quad \text{ou} \quad E[Z(x)] - E[Z(x+h)] = E[Z(x) - Z(x+h)] = 0$$

Na hipótese de estacionariedade de 2ª ordem a covariância entre os pares  $Z(x)$  e  $Z(x+h)$ , separados por um vector distância  $h$ , é estacionária e a estacionariedade da covariância implica na estacionariedade da variância

$$C[Z(x), Z(x+h)] = E[(Z(x) \cdot (Z(x+h) - m))] = E[Z(x) \cdot Z(x+h)] - m^2 \quad (3.12)$$

$$\text{Var}[Z(x)] = E[Z(x) - m]^2 = E[Z(x) \cdot Z(x+0)] - m^2 = C(0), \quad \forall x \in A$$

Se o processo é estacionário de 2ª ordem variância é constante independentemente da origem no espaço ou no tempo



**Figura 16 - Hipótese de estacionaridade de 2ª ordem [45]**

Se efeitos de 1ª e 2ª ordem são constantes, em toda a região estudada, ou seja não há tendência, o processo é considerado estacionário.

Um outro conceito básico fundamental na teoria das variáveis regionalizadas é a chamada hipótese intrínseca, a qual implica que a função (intrínseca) descreve o comportamento espacial da variável regionalizada dentro do espaço e que essa função é uma característica intrínseca da regionalização. A função intrínseca é na verdade o chamado semi-variograma.

Noutras palavras, a geoestatística assume que a distribuição das diferenças entre dois pontos amostrais é a mesma e que ela depende apenas da distância e orientação entre pontos.

A variação espacial é estacionária se ela puder ser reconhecida em todas as partes do espaço, ou seja, o variograma é o mesmo onde quer que se amostra.

Hipótese de estacionariedade intrínseca estabelece que os incrementos  $[Z(x) - Z(x + h)]$  têm esperança zero e variância somente em função de  $h$ , assim:

$$E[Z(x) - Z(x + h)] = 0, \quad \forall x \in A. \quad (3.13)$$

$$\text{Var}[Z(x) - Z(x + h)] = E\{[Z(x) - Z(x + h)]^2\} = 2\gamma(h) \quad (3.14)$$

em que  $2\gamma(h)$  é denominado de função variograma e  $g(h) = C(0) - C(h)$ . A covariância  $C(h)$  e o variograma  $2\gamma(h)$  são ferramentas equivalentes para caracterizar a dependência espacial.

Uma aplicação geoestatística deverá portanto satisfazer pelo menos uma destas hipóteses: Hipótese de estacionariedade de 2ª ordem e Hipótese intrínseca.

### 3.4.6 Variografia

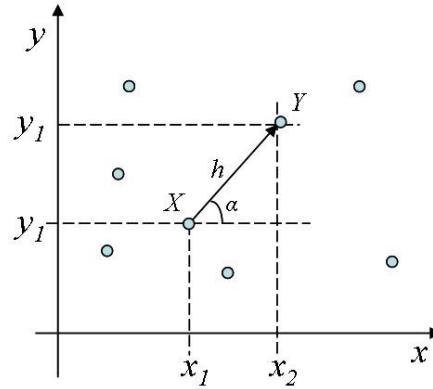
Entende-se por variografia a fase inicial de qualquer estudo geoestatístico, cujo objectivo é a construção de um modelo variográfico, que incorpore e sintetize os traços estruturais mais importantes da variável regionalizada.

Note-se, no entanto que a variografia deve ser antecedida por uma crítica cuidadosa da informação existente, que passa pela definição correcta da variável regionalizada a estudar e respectivo suporte [45].

### 3.4.6.1 Variograma

O variograma é uma ferramenta básica de suporte às técnicas de geoestatística que permite representar quantitativamente a variação de um fenómeno regionalizado no espaço. Considere duas variáveis regionalizadas,  $X$  e  $Y$ , onde  $X=Z(x)$  e  $Y=Z(x+h)$ .

Neste caso  $X$  e  $Y$  são duas amostras referentes ao mesmo atributo medido em duas posições diferentes e  $h$  o vector distância que os separa.



**Figura 17: Dependência entre duas variáveis regionalizadas**

O nível de dependência entre essas duas variáveis regionalizadas,  $X$  e  $Y$ , é representado pelo variograma ( $2\gamma(h)$ ) o qual é definido como esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados pelo vector distância  $h$ , isto é,

$$2\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} \quad (3.15)$$

$$2\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} = \text{Var}[Z(x) - Z(x+h)] \quad (3.16)$$

Através de uma amostra  $Z(x_i)$ ,  $i=1, 2, \dots, n$ , o variograma pode ser estimado por

$$2\gamma(\mathbf{h}) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2 \quad (3.17)$$

Logo analogamente o semivariograma será

$$\gamma(\mathbf{h}) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2 \quad (3.18)$$

- $2\gamma(\mathbf{h})$  - é o variograma estimado e  $\gamma(\mathbf{h})$  o semivariograma ;
- $\mathbf{h}$ : é o vector distância (módulo e direcção) entre pares de observação;

- $N(h)$  - é o número de pares de valores medidos,  $z(x_i)$  e  $z(x_i+h)$ , separados por uma distância  $h$ ;
- $Z(x_i)$  e  $Z(x_i + h)$  - são valores recolhidos nos pontos  $x_i$  e  $x_i+h$

Na construção do semivariograma, são necessárias para o cálculo de  $\gamma(h)$  um mínimo de 30 pares (regra prática), o que pode ser conseguido se for escolhido como maior  $\gamma(h)$ , a metade da maior distância existente entre os pontos. [40]

#### **3.4.6.2 Parâmetros do semivariograma**

Os semivariogramas expressam o comportamento espacial da variável regionalizada e mostram:

- Alcance ou Amplitude (a): O tamanho da zona de influência em torno de uma amostra. Assim toda a amostra cuja distância ao ponto a ser estimado for menor ou igual ao alcance, fornece informações sobre o ponto. A amplitude é a distância que separa o campo estruturado (amostras correlacionadas) do campo aleatório (amostras independentes)
- Patamar (C): É o valor da variância no qual o variograma se estabiliza. Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ( $\text{Var}[Z(x) - Z(x+h)]$ ) torna-se invariante com a distância.
- Efeito Pepita ( $C_0$ ): O valor da função variograma na origem ( $h=0$ ). Teoricamente este valor deveria ser zero pois duas amostras recolhidas no mesmo ponto deveriam apresentar o mesmo valor. O efeito pepita é atribuído a erros de medição ou ao facto de que os dados não foram recolhidos em intervalos suficientemente pequenos para mostrar o comportamento espacial subjacente do fenómeno em estudo.
- Variância espacial: Diferença entre a variância e o efeito pepita

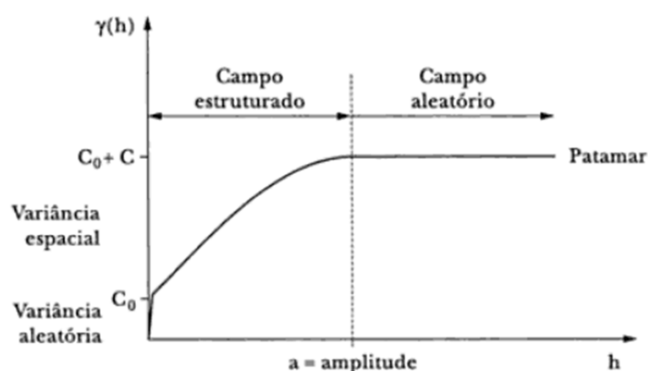


Figura 18- Semivariograma experimental [38]

### 3.4.6.3 Cálculo do semivariograma a partir de amostras regularmente espaçadas

Seja uma variável regionalizada  $x$  amostrada em diversos pontos  $i$  regularmente distribuídos por certa região. O grau de relação entre pontos numa certa direcção pode ser expresso pela covariância e, embora a covariância exista entre todas as distâncias possíveis ao longo de  $h$ , pode ser estipulado que somente sejam considerados valores entre pontos regularmente espaçados por múltiplos inteiros de  $\Delta h$ .

Assim sendo considera-se o conjunto de amostras regularmente espaçadas, em duas dimensões, conforme apresentado na Figura 19. Em Clark 1979 é apresentada a construção de um semivariograma a partir de uma rede regular, com espaçamento entre os pontos de 100 pés. Trata-se de um depósito estratiforme de ferro com valores em percentagem por peso.

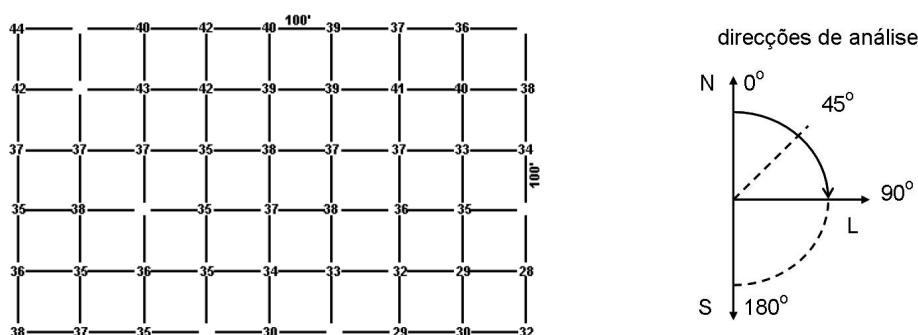


Figura 19 – Conjunto de amostras regularmente espaçadas[50]

Para determinar o semivariograma experimental, por exemplo, na direcção de 90° o cálculo de  $\gamma(h)$  é repetido para todos os intervalos de  $h$

$$\gamma(\mathbf{h}) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2 \quad (3.19)$$

Para  $h=100$  pés.

$$\begin{aligned} \gamma^*(100) = & [(40 - 42)^2 + (42 - 40)^2 + (40 - 39)^2 + (39 - 37)^2 + (37 - 36)^2 + (43 - 42)^2 + (42 - 39)^2 \\ & + (39 - 39)^2 + (39 - 41)^2 + (41 - 40)^2 + (40 - 38)^2 + (37 - 37)^2 + (37 - 37)^2 + (37 - 35)^2 + (35 - 38)^2 + \\ & (38 - 37)^2 + (37 - 37)^2 + (37 - 33)^2 + (33 - 34)^2 + (35 - 38)^2 + (35 - 37)^2 + (37 - 36)^2 + (36 - 36)^2 + (36 \\ & - 35)^2 + (36 - 35)^2 + (35 - 36)^2 + (36 - 35)^2 + (35 - 34)^2 + (34 - 33)^2 + (33 - 32)^2 + (32 - 29)^2 + (29 - \\ & 28)^2 + (38 - 37)^2 + (37 - 35)^2 + (29 - 30)^2 + (30 - 32)^2] / [2 \times 36] = 1,46 \end{aligned}$$

Para  $h=200$  pés.

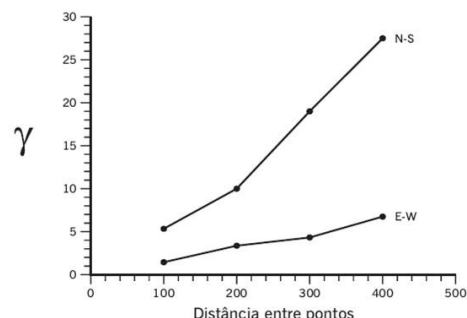
$$\begin{aligned} \gamma^*(200) = & [(44 - 40)^2 + (40 - 40)^2 + (42 - 39)^2 + (40 - 37)^2 + (39 - 36)^2 + (42 - 43)^2 + (43 - 39)^2 + \\ & (42 - 39)^2 + (39 - 41)^2 + (39 - 40)^2 + (41 - 38)^2 + (37 - 37)^2 + (37 - 35)^2 + (37 - 38)^2 + (35 - 37)^2 + (38 - \\ & 37)^2 + (37 - 33)^2 + (37 - 34)^2 + (38 - 35)^2 + (35 - 36)^2 + (37 - 36)^2 + (36 - 35)^2 + (36 - 36)^2 + (35 - 35)^2 \\ & + (36 - 34)^2 + (35 - 33)^2 + (34 - 32)^2 + (33 - 29)^2 + (32 - 28)^2 + (38 - 35)^2 + (35 - 30)^2 + (30 - 29)^2 + (29 \\ & - 32)^2] / [2 \times 33] = 3,30 \end{aligned}$$

O mesmo procedimento é feito para o cálculo do semivariograma na direcção Norte-Sul. Os resultados estão expressos na tabela 4. Estes resultados permitem a construção dos semivariogramas experimentais, nas duas direcções consideradas, e o que se pode perceber é que há uma distinta diferença na estrutura dos dados, ou seja, a presença de uma anisotropia pois os semivariogramas mostram diferentes comportamentos para diferentes direcções de linhas de amostragem e de estudo da variável.



**Tabela 4: Valores da função semivariograma calculados para as direcções Este- Oeste e Norte – Sul e semivariogramas experimentais [40]**

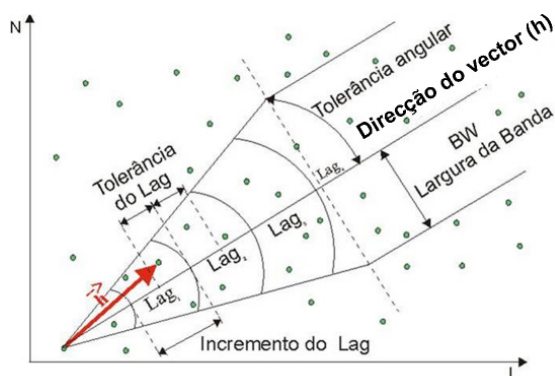
Direcção	Distância	Semivariograma	Nº Pares
Este – Oeste (90°)	100	1,46	36
	200	3,30	33
	300	4,31	27
	400	6,70	23
Norte-Sul	100	5,35	36
	200	9,87	27
	300	18,88	21



#### 3.4.6.4 Cálculo do semivariograma a partir de amostras irregularmente espaçadas

Para dados que estão irregularmente distribuídos no espaço bidimensional não é possível, em princípio, encontrar pares de amostras suficientes com exactamente o mesmo espaçamento  $\Delta h$  para o cálculo numa determinada direcção, como feito com dados dispostos em malha regular. [40]

Para contornar essa situação definem-se quatro parâmetros adicionais: Incremento de *lag*, tolerância de *lag*, tolerância angular e largura de banda como se pode ver na figura 20.



**Figura 20 - Parâmetros de um semivariograma para amostras irregularmente espaçadas [45]**

Tomemos como exemplo a  $Lag_2$  (*Lag* refere-se a uma distância pré-definida, a qual é utilizada no cálculo do semivariograma) da figura acima.

Suponha um incremento de *Lag* igual a 100 metros com tolerância de 50 metros. Considere ainda a direcção de medida 45° com tolerância angular 22.5°. Então, qualquer par

de observações cuja distância está compreendida entre 150 m e 250 m e  $22.5^\circ$  e  $67.5^\circ$  será incluído no cálculo do semivariograma de  $Lag_2$ . Este processo repete-se para todos as  $Lag$ 's. Ainda com referência na Figura 20, a largura de banda ( $BW$ ) refere-se a um valor de ajuste a partir do qual se restringe o número de pares de observações para o cálculo do semivariograma.

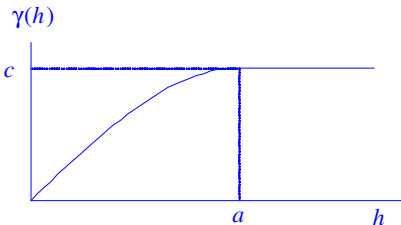
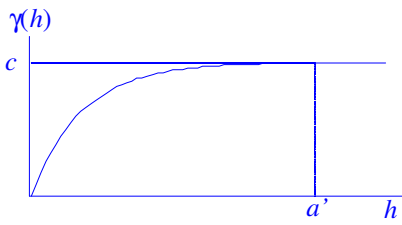
Ao cálculo deste semivariograma dá-se o nome de semivariograma direccional pois é necessário especificar qual a direcção em que queremos fazer esta análise.

No caso de se pretender o semivariograma com informação em todas as direcções usa-se o semivariograma omnidireccional.

### 3.4.6.5 Ajuste do modelo teórico

De posse do semivariograma experimental é necessário ajustá-lo a um modelo teórico, ou seja, a uma função que definirá os parâmetros do semivariograma, “efeito pepita”, “alcance” e “patamar”. Entre os modelos disponíveis estes são os mais utilizados:

**Tabela 5- Modelos teóricos de ajuste**

<i>Esférico</i>	<i>Exponencial</i>
$\gamma(h) = \begin{cases} c \left( 1.5 \frac{h}{a} - 0.5 \frac{h^3}{a^3} \right) & h \leq a \\ c & h > a \end{cases}$ 	$\gamma(h) = c [1 - e^{-(h/a)}]$ $a' = 3a \rightarrow \gamma(a') = 0.95c$ 

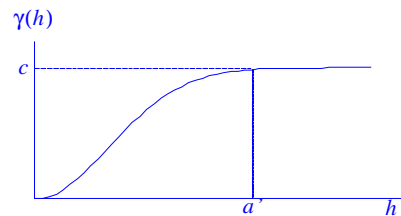
---

**Gaussiano**


---

$$\gamma(h) = c \left[ 1 - e^{-(h/a)^2} \right]$$

$$a' = \sqrt{3} a \rightarrow \gamma(a') = 0.95 c$$



A modelização, ou seja, o ajuste de um variograma experimental a uma função é um passo fundamental na análise variográfica, sendo um processo que envolve várias tentativas e na qual a experiência pesa muito.

Pode-se optar por um ajuste manual por comparação visual, mais sujeito a erros, ou, com o auxílio de algoritmos, para ajustes automáticos como apresentado, entre outros, em *Pannatier (1996)*.

Os modelos aqui apresentados são considerados modelos básicos, tendo sido denominados de modelos isotrópicos por *Isaaks e Srivastava (1989)*. Estão divididos em dois tipos: modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na geoestatística como modelos transitivos. Alguns dos modelos transitivos atingem o patamar (C) assintoticamente. Para tais modelos, o alcance (a) é arbitrariamente definido como a distância correspondente a 95% do patamar. Modelos do segundo tipo não atingem o patamar, e continuam aumentando enquanto a distância aumenta.

Com a variografia pretende-se o cálculo e interpretação dos variogramas experimentais ao longo de várias direcções e modelização (ajuste) de um modelo teórico, válido para todo o campo, que sintetize o comportamento espacial da variável.

#### 3.4.6.6 Isotropia

Quando a variabilidade espacial de um fenómeno em estudo é a mesma em todas as direcções, diz-se que o fenómeno é isotrópico (Figura 21).

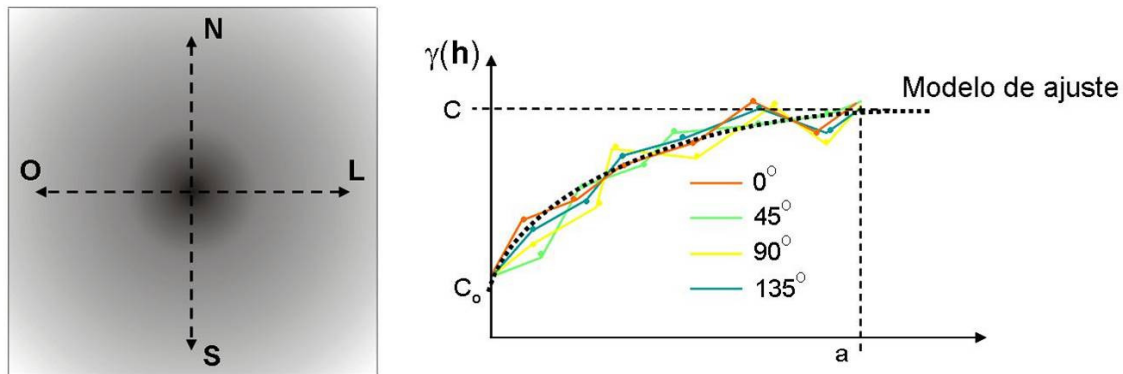


Figura 21 – Fenómeno isotrópico [40]

Neste caso, um único modelo é suficiente para descrever a variabilidade espacial do fenómeno em estudo.

Por outro lado e depois de se modelizar o semivariograma direccionado e se concluir que a variabilidade espacial não é mesma em todas as direcções diz-se que o fenómeno é anisotrópico. A análise da anisotropia tem como objectivo detectar as direcções de maior e menor continuidade espacial do fenómeno investigado. A anisotropia pode ser facilmente constatada através da observação dos semivariogramas obtidos para diferentes direcções

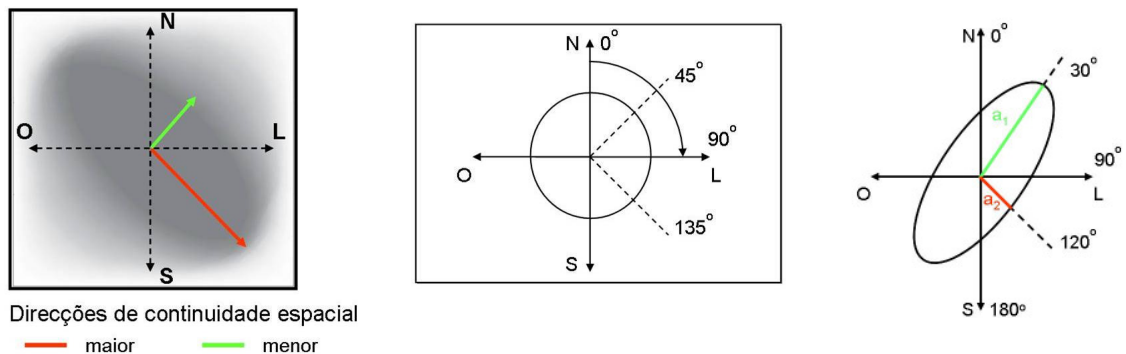


Figura 22 - Fenómeno anisotrópico [40]

O modo directo de visualizar e calcular os parâmetros (factor e ângulo) da anisotropia é através de um esboço gráfico de uma elipse (ou diagrama de rosa). Aqui temos que ter em atenção dois parâmetros: Factor e ângulo de anisotropia. Sendo que o factor de anisotropia é a relação entre a maior e a menor direcção de continuidade espacial e o ângulo de anisotropia o ângulo referente ao maior alcance (exemplo 30°).

$$F_a = \frac{a_2}{a_1} \quad (3.20)$$

### 3.4.6.7 Tipos de anisotropia

Há anisotropia, quando os semivariogramas mostram diferentes comportamentos para diferentes direcções de linhas de amostragem e de estudo da variável. Neste caso a anisotropia pode ser geométrica quando o alcance varia de acordo com as diversas direcções consideradas, mantendo constante o patamar e zonal quando o alcance permanece constante e o patamar varia conforme for modificada a direcção.

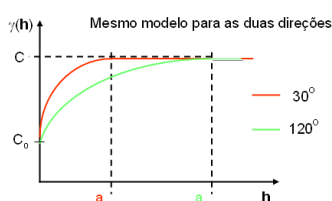


Figura 23 - Anisotropia geométrica

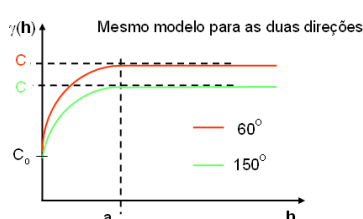


Figura 24 - Anisotropia zonal [40]

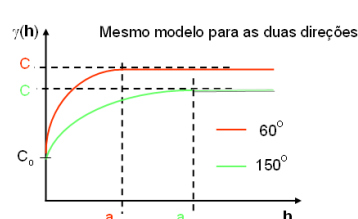


Figura 25 - Anisotropia combinada

Para a estimativa do semivariograma experimental, não se tendo certeza se o fenómeno sob estudo é isotrópico ou anisotrópico no espaço bidimensional, inicialmente consideram-se quatro direcções, E-W, N-S, NE-SW e NW-SE, com tolerância angular de 45°.

Melhores estimativas são obtidas quando os modelos são baseados em semivariogramas experimentais que apresentam a menor razão “efeito pepita/patamar” e, também, o maior alcance.

Uma outra consideração importante a ser feita é determinar o grau de aleatoriedade presente nos dados pela fórmula  $E = C_0 / C$  (Guerra 1988):

- $E < 0,15$ : componente aleatória pequena;
- $0,15 \leq E \leq 0,30$ : componente aleatória significativa;
- $E > 0,30$ : componente aleatória muito significativa. [40]

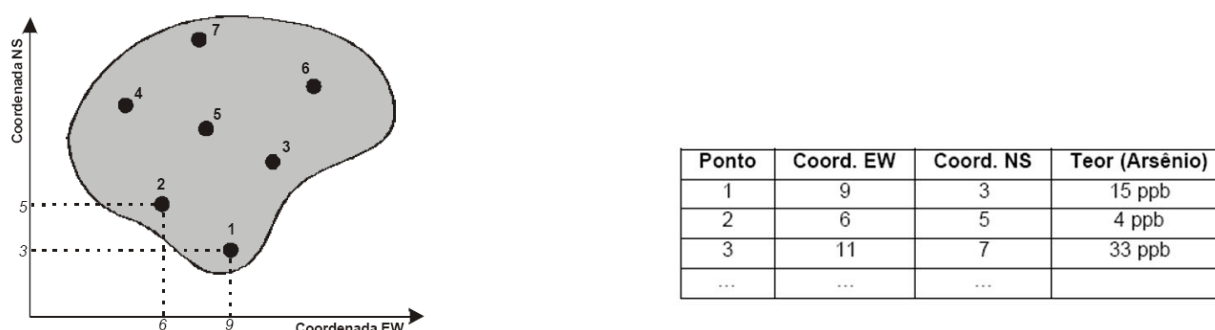
## 3.5 Estimação geostatística

### 3.5.1 Introdução

A geoestatística tem por objecto a caracterização da dispersão espacial e espacio-temporal das grandezas que definem a quantidade e a qualidade de recursos naturais, tais como as florestas, recursos hidrológicos e ecológicos. [18]

A apresentação gráfica da dispersão espacial destas grandezas pode ser feita através de mapa de contorno com a apresentação através de curvas de isovalores chamadas isolinhas podendo a apresentação dos resultados ter outro tipo de apresentação gráfica (Superfície 3D, imagem sombreada) .

O primeiro passo em qualquer estudo espacial é a definição do delineamento experimental, que envolve, entre outros procedimentos, a escolha da técnica de recolha de amostras e também da malha de amostragem.



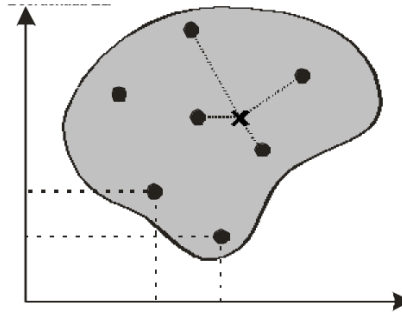
**Figura 26 - Pontos de amostragem numa área de estudo [51]**

O número sobre cada ponto é a sua identificação e os números nos eixos das ordenadas e abcissas são as coordenadas dos pontos no espaço.

Nesses pontos são medidas as variáveis em estudo como por exemplo o teor de concentração de um elemento químico específico.

A partir das coordenadas do ponto e dos valores assumidos pela variável nesses pontos, pode-se obter um mapa da distribuição espacial para cada variável, utilizando-se um algoritmo de interpolação.

Interpolarm não é nada mais do que estimar o valor da variável em estudo num ponto não amostrado. Na figura 27 a cruz representa o ponto que se pretende estimar o valor.



**Figura 27 - Pontos de amostragem e de estimativa numa área de estudo [51]**

Basicamente os algoritmos de interpolação funcionam fornecendo pesos aos pontos amostrados na predição diferindo, entre eles, a maneira de atribuir os pesos às amostras.

A base de muitos métodos de interpolação é a mesma. O valor de um ponto a ser estimado ( $Z^*$ ) é o somatório do produto entre o valor da variável de um ponto conhecido  $Z(x_i)$  por um peso calculado ( $\lambda_i$ ) para os pontos  $i$ , variando de 1 até  $N$ , onde  $N$  representa o número total de pontos considerados.

$$Z^* = \sum_{i=1}^n \lambda_i \cdot Z(x_i) \quad (3.21)$$

Para isso, pode-se utilizar os pontos já conhecidos, no caso da figura anterior, pontos 3, 5, 6 e 7. Um primeiro valor seria a média aritmética dos pontos mais próximos, assim os pesos dos pontos seriam iguais. Para 4 pontos utilizados ( $N=4$ ), na forma geral, tem-se:

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{4} \quad (3.22)$$

Supondo que os valores da variável em estudo ( $Z$ ) nos pontos 3, 5, 6 e 7 sejam  $Z_3=300$ ,  $Z_5=100$ ,  $Z_6=200$  e  $Z_7=500$ , o valor de  $Z$  no ponto predito ( $Z^*$ ) será:

$$Z^* = \sum_{i=1}^n \lambda_i \cdot Z(x_i) = \frac{1}{4} 300 + \frac{1}{4} 100 + \frac{1}{4} 200 + \frac{1}{4} 500 = 275 \quad (3.23)$$

Na verdade, é até intuitivo que, quando trabalhando no espaço, pontos conhecidos mais próximos aos pontos a serem preditos devam ter um peso maior, pois sua influência é maior. Assim, poder-se-ia utilizar um outro algoritmo para calcular os pesos, baseado no Inverso da Distância, que estabelece que quanto menor for a distância entre o ponto conhecido e o predito, maior será o peso. Supondo que essa distância seja  $d_3=2$ ,  $d_5=2$ ,  $d_6=4$  e  $d_7=5$ , tem-se:

$$Z^* = \sum_{i=1}^n \lambda_i \cdot Z(x_i) = \frac{1}{2} 300 + \frac{1}{2} 100 + \frac{1}{4} 200 + \frac{1}{5} 500 = 350 \quad (3.24)$$

Dessa forma, pode-se ver que o valor predito depende directamente do método de interpolação escolhido: no exemplo, para a Média Aritmética o valor obtido é  $Z^*=300$  e para o Inverso da Distância é  $Z^*=350$ .

Qual é, porém, o valor real? A única maneira de saber é medindo a variável no ponto que foi predito, sabendo ainda que poderá haver um erro associado à medida, dado pela calibração do equipamento utilizado, da técnica de medida e outros erros associados. Assim, a questão passa a ser qual é o valor mais próximo ao real? Neste caso a resposta é bem mais difícil.

Existe um número significativo de métodos de interpolação que podem fornecer resultados divergentes, mas cada um vai ter uma resposta para cada variável em estudo

Em relação aos ponderadores ( $\lambda_i$ ) pretende-se que eles resumam dois efeitos extremamente importantes num processo de inferência espacial: em primeiro lugar, devem reflectir a maior ou menor “proximidade” estrutural das amostras ( $Z^*$ ) em relação ao ponto a estimar  $Z(x_i)$  e ao mesmo tempo, devem ter o efeito desagregador de eventuais agrupamentos preferências (*clusters*) de amostras, de modo a evitar o enviesamento causado por estes. [18]

O que se conhece destes fenómenos espaciais resume-se normalmente a um conjunto discreto e limitado de amostras ou observações. O problema básico da geoestatística é o da caracterização da distribuição espacial daquelas grandezas e o da avaliação de medidas de incerteza, tendo em conta a maior ou menor variabilidade do fenómeno espacial, a qualidade das amostras e observações o tipo de modelo geoestatístico escolhido e o maior ou menor grau de conhecimento que se tem do fenómeno. [18]

As variáveis que medem fenómenos naturais – geologia, geoquímica, cartografia, avaliação de recursos florestais, apresentam em comum o facto de dependerem fortemente da localização espacial, sendo por isso chamadas de variáveis regionalizadas. [37]

Nas ciências da terra e do ambiente, os métodos clássicos de inferência estatística espacial de variáveis características de um recurso natural, desde o polígono de influência, o inverso do quadrado das distâncias e os ajustamentos polinomiais, tem, na sua maioria, duas características comuns: trata-se de métodos deterministas sem qualquer medida de incerteza associada e não tomarem em linha de conta medidas de continuidade estrutural como os variogramas ou covariâncias do fenómeno natural em estudo [18].



### 3.5.2 Estimador Linear Geoestatístico: *Krigagem*

O problema de encontrar o melhor estimador do valor real tomado por uma variável regionalizada num dado espaço conhecendo alguns valores através de amostras colhidas foi resolvido por *G. Matheron* e que em homenagem a *Daniel G. Krige* deu o nome de *Krigagem*

O método proposto consiste essencialmente em procurar um estimador, combinação linear de toda a informação disponível, cujos coeficientes (ponderadores associados aos diversos valores amostrados) obedecem a duas condições, uma de universalidade (não enviesamento) e outra de optabilidade.

Inicialmente, o método de *krigagem* foi desenvolvido para solucionar problemas de mapeamentos geológicos, mas seu uso expandiu-se com sucesso no mapeamento de solos (*Burgess e Webster*, 1980a,b), mapeamento hidrológico (*Kitanidis e Vomvoris*, 1983), mapeamento atmosférico (*Lajaunie*, 1984) e outros campos correlacionados [52].

*Krigagem* está baseado no conceito de variável regionalizada. A diferença entre a *krigagem* e outros métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras. No caso de interpolação linear simples, por exemplo, os pesos são todos iguais a  $1/N$  ( $N$  = número de amostras); na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados. [47]

A *krigagem* usa informações a partir do semivariograma para encontrar os pesos óptimos a serem associados às amostras que irão estimar um ponto, uma área ou um bloco. Como o semivariograma é uma função da distância entre locais de amostragens, mantendo o mesmo número de amostras, os pesos são diferentes de acordo com o seu arranjo geográfico. O uso do semivariograma para a estimativa por *krigagem* não exige que os dados tenham distribuição normal, mas a presença de distribuição assimétrica, com muitos valores anómalos, deve ser considerada, pois a *krigagem* é um estimador linear.[40]

#### 3.5.2.1 Tipos de *Krigagem*

Dois métodos de *Krigagem* gerais e amplamente usados são a *Krigagem Ordinária* e *Universal*. A *krigagem Universal* assume que há uma tendência nos dados. Por exemplo, pode-se saber que há a prevalência de vento ou uma encosta de declive suave ao longo da sua

área de estudo. A *krigagem* ordinária assume que não há tendência nos dados, que deve ser sua suposição padrão de operação.

### 3.5.2.2 *Krigagem Simples*

$$Z_{x_0}^* = \lambda_0 + \sum_{i=1}^n \lambda_i \cdot Z(x_i) \quad (3.25)$$

O método de *krigagem* simples supõe que a média ( $m$ ) é conhecida e constante *a priori*, então:

$$E[Z(x_0)] = E[Z(x_i)] = m \quad (3.26)$$

Logo

$$Z_{x_0}^* = m + \sum_{i=1}^n \lambda_i \cdot [Z(x_i) - m] \quad (3.27)$$

### 3.5.2.3 *Krigagem Ordinária*

Diferente da *krigagem* simples, a *krigagem* ordinária não requer o prévio conhecimento da média  $m$ . Neste caso, para que a igualdade da Equação 2.41 seja satisfeita é necessário que:

$$\lambda_0 = 0 \text{ e } \sum_{i=1}^n \lambda_i = 1 \quad (3.28)$$

Portanto, o estimador de *krigagem* ordinária é

$$Z_{x_0}^* = \sum_{i=1}^n \lambda_i \cdot [Z(x_i)] \quad (3.29)$$

## 3.5.3 Sistema de equações de *krigagem*

A informação disponível consiste geralmente numa informação experimental e numa informação estrutural consubstanciada no modelo da função variograma que reflecte as características da variabilidade espacial da variável regionalizada. [37]

A informação experimental é constituída por um conjunto discreto de  $N$  valores  $Z_i$  (com  $i=1, N$ ). Estes valores são, ou valores pontuais  $Z(x_i)$  ou valores médios  $Z_i$ , definidos em suportes de dimensão  $V_i$  (volume, superfície ou comprimento).

O estimador de krigagem  $Z^*$  tomará então a forma :  $Z^* = \sum_{i=1}^n \lambda_i \cdot Z(x_i)$  em que  $\lambda_i$  são os ponderadores a afectar à informação experimental ou seja são os pesos atribuídos a cada amostra  $i$ .

Se a soma dos pesos for igual a 1 e não ocorrer tendência local dos valores, esse estimador é o melhor e não tendencioso, pois a partir dos pesos atribuídos a cada amostra, minimiza a estimativa da variância. O objectivo da *krigagem* é procurar pelo conjunto óptimo de ponderadores de modo que a variância do erro de estimativa seja a menor possível.

O sistema de *krigagem* ordinária pode exprimir-se na forma matricial:

$$[M] = [\lambda] \cdot [G] \quad (3.30)$$

logo

$$[\lambda] = [G]^{-1} \cdot [M] \quad (3.31)$$

em que a matriz do 1º membro ou matriz de *krigagem* [G] se escreve

$$G = \begin{bmatrix} \bar{\gamma}(x_1, x_1) & \bar{\gamma}(x_1, x_2) & \cdots & \bar{\gamma}(x_1, x_n) & 1 \\ \bar{\gamma}(x_2, x_1) & \bar{\gamma}(x_2, x_2) & \cdots & \bar{\gamma}(x_2, x_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \bar{\gamma}(x_n, x_1) & \bar{\gamma}(x_n, x_2) & \cdots & \bar{\gamma}(x_n, x_n) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (3.32)$$

A matriz dos ponderadores  $[\lambda]$  de *krigagem* e a matriz 2º membro [M] são duas matrizes-colunas

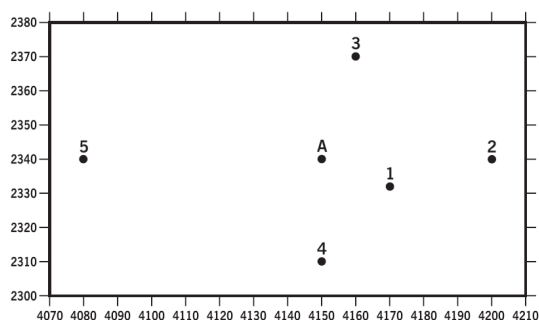
$$M = \begin{bmatrix} \bar{\gamma}(x_1, x_0) \\ \bar{\gamma}(x_2, x_0) \\ \vdots \\ \bar{\gamma}(x_n, x_0) \\ 1 \end{bmatrix} \quad \text{e} \quad \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} \quad (3.33)$$

A matriz [G] contém os valores obtidos no semivariograma referentes às distâncias entre as amostras estimadoras; a matriz coluna [M] contém os valores obtidos no semivariograma

referentes às distâncias entre cada amostra e o ponto ( $x_0$ ) a ser estimado e a matriz  $[\lambda]$  contém os ponderadores a serem calculados.

### 3.5.4 Exemplo Prático - Krigagem simples

Consideremos o espaço amostral da Figura 28.



Pontos	X	Y	$U_0$	Distâncias
A	4150	2340	?	00,00
1	4170	2332	400	21,54
2	4200	2340	380	50,00
3	4160	2370	450	31,62
4	4150	2310	280	30,00
5	4080	2340	320	70,00

Figura 28 – Localização espacial de pontos de amostragem [40]

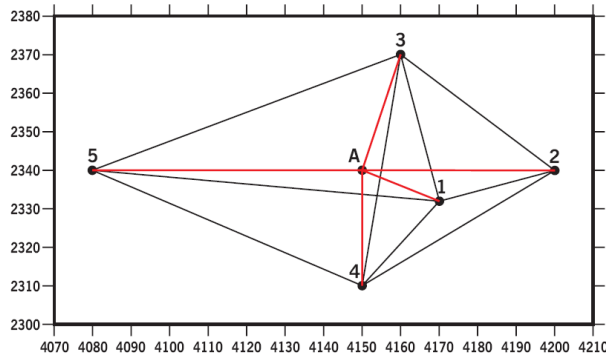
Na *krigagem* há necessidade de encontrar, para além das distâncias euclidianas, preliminarmente, a variabilidade espacial dos dados o que é conseguido por uma análise variográfica. Noutras palavras, antes da aplicação da *krigagem* é preciso calcular um semivariograma experimental e em seguida modelá-lo.

Tendo isso sido feito encontrou-se um modelo esférico, com os seguintes parâmetros: amplitude de influência ( $a$ ), 100 pés; valor da soleira ( $C$ ), 700 ppm e um efeito pepita ( $C_0$ ) da ordem de 100 ppm (Fig. 28).

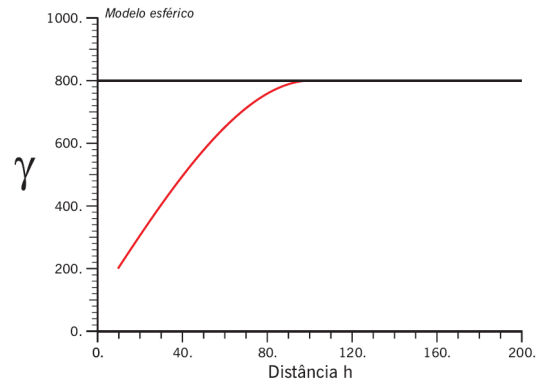
$$\gamma(h) = C \left[ \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right] + C_0 \quad (3.34)$$

$$\gamma(21,54) = 700 \left[ \frac{3}{2} \left( \frac{21,54}{100} \right) - \frac{1}{2} \left( \frac{21,54}{100} \right)^3 \right] + 100 = 322,7 \quad (3.35)$$

Na *krigagem* para a estimativa de A são necessários não apenas os valores das distâncias de cada um dos 5 pontos em relação a A, mas também as distâncias entre os pontos.



**Figura 29 – Inter-relações entre amostras e o ponto a ser estimado para cálculo das equações de krigagem. [40]**



**Figura 30 – Modelo de variograma para os dados da figura 28.**

Na posse desses valores, pode-se organizar o sistema de equações normais para o cálculo dos pesos  $\lambda_i$ .

$$\begin{bmatrix} 0 & 415,5 & 491,4 & 403,0 & 790,5 & 1 \\ 415,5 & 0 & 581,3 & 642,9 & 800,0 & 1 \\ 491,4 & 581,3 & 0 & 659,9 & 778,8 & 1 \\ 403,0 & 642,9 & 659,9 & 0 & 745,1 & 1 \\ 790,5 & 800,0 & 778,8 & 745,1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{bmatrix} = \begin{bmatrix} 322,7 \\ 581,3 \\ 420,9 \\ 405,5 \\ 714,9 \\ 1 \end{bmatrix} \quad (3.36)$$

$$[G][\lambda] = [M] \quad (3.37)$$

$$G^{-1} = \begin{bmatrix} -0,002246 & 0,000802 & 0,000535 & 0,000911 & -0,000002 & 0,038234 \\ 0,000802 & -0,001476 & 0,000343 & 0,000085 & 0,000246 & 0,216190 \\ 0,000535 & 0,000343 & -0,001314 & 0,000148 & 0,000288 & 0,215872 \\ 0,000911 & 0,000148 & 0,000154 & -0,001501 & 0,000357 & 0,214851 \\ -0,000002 & 0,000288 & 0,000278 & 0,000357 & -0,000889 & 0,314853 \\ 0,038234 & 0,215872 & 0,223193 & 0,214851 & 0,314853 & -531,3827 \end{bmatrix}$$

$$[\lambda] = [G]^{-1}[M] = \begin{bmatrix} 0,3760 \\ -0,0347 \\ 0,3017 \\ 0,2668 \\ 0,0901 \\ 10,0816 \end{bmatrix} \quad (3.38)$$

Para o cálculo do teor de A

$$A = (0,3760 \cdot 400) + (-0,0347 \cdot 380) + (0,3017 \cdot 450) + (0,2668 \cdot 280) + (0,0901 \cdot 320) = 376,55 \text{ ppm}$$

A variância associada à estimativa pela *krigagem* é:

$$S^2 = 0,376(322,7) - 0,0347(581,3) + 0,3017(420,9) + 0,2668(405,5) + 0,0901(714,9) + 10,0816 = 410,8867$$

$$S = \sqrt{410,8867} \Leftrightarrow S = 20,27034 \quad (3.39)$$

## **4 Sistema Desenvolvido**

### **4.1 Solução para um veículo autónomo de superfície**

O estudo de sistemas físicos como oceanos, rios ou lagos, caracterizando as suas propriedades, compreendendo os processos que neles ocorrem e as interacções com os sistemas vizinhos, e previsão do seu comportamento dinâmico são matéria de estudo de diferentes ciências.

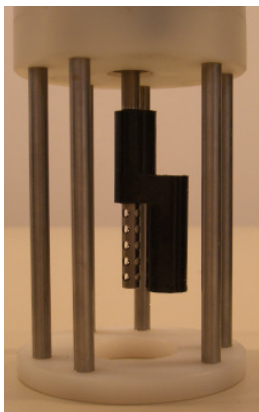
Mas a observação do oceano apresenta dificuldades acrescidas relativamente à de outros sistemas físicos por se tratar de um meio extenso, de difícil e oneroso acesso, que é praticamente opaco à radiação electromagnética, e que obriga os instrumentos de medição a suportar fortes pressões e uma corrosão intensa [43] .

Os métodos de observação utilizados têm evoluído ao longo do tempo em função do desenvolvimento tecnológico que se vai atingindo. O rápido progresso da electrónica veio modificar drasticamente o modo de observar o oceano, permitindo descobrir muitos aspectos até aí ignorados e a existência de uma enorme diversidade de escalas espaciais e temporais.

As propriedades físicas básicas da água do mar são a temperatura e a salinidade. A partir destes parâmetros é possível calcular diversas outras propriedades derivadas, de entre as quais sobressai a densidade da água, pela sua importância no estudo da estática e da dinâmica do oceano. A profundidade permite-nos a monitorização de perfilagem das grandezas.

O CTD é um instrumento de medida multiparâmetro para monitorização estacionária ou de perfilagem de três grandezas físicas principais: Pressão (Profundidade), Temperatura e Condutividade.

Devido às particularidades da aplicação a que se destinam é um instrumento que efectua medições com grande precisão e incorpora sensores de temperatura, condutividade e pressão.



**Figura 31 - CTD**



**Figura 32 - Veículo autónomo de superfícies - ROAZ**

A solução desenvolvida consiste em usar um CTD que futuramente irá ser integrado no veículo autónomo ROAZ ([http://www.lsa.isep.ipp.pt/pages/roaz\\_roaz.html](http://www.lsa.isep.ipp.pt/pages/roaz_roaz.html)) para monitorização marítima e fluvial conforme se pode ver na figura 32.

#### **4.1.1 Grandezas físicas**

##### **4.1.1.1 Temperatura**

A medição da temperatura da água do mar/rio baseia-se na variação da resistência eléctrica de um metal por efeito da temperatura.

Na maioria dos metais a resistência aumenta com o aumento da temperatura, tendo como principais características serem dispositivos lineares e contemplarem grandes faixas de operação. O metal mais utilizado é a platina pois possui uma alta estabilidade e uma resposta rápida.

A sonda utilizada nas medições com o CTD incorpora uma PT-100, ou seja uma resistência de platina que a 0 °C possui uma resistência de 100  $\Omega$ . Os limites de entrada entre os quais o sensor pode variar vão desde – 2 °C a 36 °C.

O tempo que o sensor de temperatura leva a entrar em equilíbrio térmico com o ambiente onde está mergulhado chama-se “constante de tempo” do sensor; a constante de tempo da PT 100 é menor que 150 ms.



#### **4.1.1.2 Condutividade**

A água pura é má condutora de electricidade, embora não seja um isolante perfeito. A condutividade (o inverso da resistividade) é normalmente usada para medir a quantidade de sal na água – um importante indicador da qualidade da água e indica a facilidade de conduzir corrente eléctrica de um material ou substância.

A concentração de iões dissolvidos (teor de sal) é o maior factor que determina a condutividade. A mobilidade desses iões também é importante – a mobilidade (e a condutividade) aumentam com a temperatura e pressão. A mobilidade e a condutividade diminuem quando a temperatura diminui.

A condutividade eléctrica da água é actualmente expressa em unidades como mS/cm (mili-Siemens por centímetro). É comum, no entanto, encontrar outras unidades de medida tais como o mho/cm, ou encontrar submúltiplos tais como  $\mu\text{S}$  (microSiemens). Para as maiorias das aplicações práticas em oceanografia, as mudanças de condutividade são seguidas por mudanças de temperatura.

O sensor de temperatura tipo PT-100, fornece uma resposta imediata as variações de temperatura do processo, é usada para compensar as variações de medição ocasionadas pela temperatura.

O sensor de condutividade utiliza células do tipo indutivo em que existem dois enrolamentos toroidais e coaxiais imersos na amostra de água contida numa célula de dimensões fixas. Uma corrente alternada ao passar no enrolamento primário induz uma força electromotriz e uma corrente no secundário, as quais são proporcionais à condutividade da amostra de água. O sensor utilizado permite medir condutividades até 70 mS/cm.

#### **4.1.1.3 Pressão**

Uma sonda CTD permite obter perfis de condutividade, temperatura e salinidade ao longo da coluna de água. A profundidade em metros é obtida indirectamente através do sensor de pressão.

Na realidade, a qualquer profundidade o CTD está submetido à pressão absoluta que é igual à soma da pressão atmosférica existente à superfície da água com a pressão exercida pela água conforme podemos ver na equação 4.1.

$$p = p_0 + \rho \cdot g \cdot h \quad (4.1)$$

Na superfície da água a pressão atmosférica é de  $p_0 = 1,01 \times 10^5$  Pa.

Na água, existe uma relação numérica de correspondência entre a mudança de pressão em decibar e a variação da profundidade relativa à superfície do mar em metros. Especificamente, o aumento de 1 decibar ocorre a cada 1,019716 m aumento de profundidade em relação à superfície.



**Figura 33 - Sensor de pressão**

Quando uma força é aplicada a uma estrutura, o comprimento desta muda. O transdutor que mede a deformação é conhecido por extensómetro (*strain gauge*) e a distorção causada pela aplicação de uma força resulta na mudança da sua resistência eléctrica conforme se pode ver na figura 33. O sensor usado utiliza este princípio e converte em tensão sinais até 25 bar, tendo um *offset* (valor de saída quando a entrada é zero) de 20 mbar

#### **4.1.1.4 Salinidade**

A salinidade é uma medida da quantidade de sais dissolvidos na água, e o mais comum é ser calculada a partir da condutividade da solução. Como regra geral, quanto maior a concentração de sais numa solução, melhor é a sua capacidade para conduzir electricidade.

A razão entre as condutividades da amostra da água do mar e da água padrão permite, através da aplicação de um polinómio estabelecido em 1978 (*UNESCO*, 1981), calcular a salinidade da amostra. Para transformar a razão de condutividade em salinidade também podem ser usadas as tabelas internacionais da *UNESCO*.

A concentração de sal que existe na água do mar é relativamente pequena. Em média, podemos falar de um valor próximo de 35g por cada quilograma de água do mar, o que corresponde aproximadamente ao valor 35 em salinidade. No entanto, esta pequena quantidade de sal leva a grandes diferenças nas propriedades físicas da água do mar relativamente às da água pura. Um exemplo concreto é o abaixamento da temperatura de

fusão da água quando se lhe juntam sais (a água do mar ainda se conserva no estado líquido mesmo quando a temperatura atinge cerca de  $-2\text{ }^{\circ}\text{C}$ ).

#### **4.1.2 CTD**

O CTD utilizado é o SW100 da globo que é composto por diferentes componentes modulares, que podem ser combinados em diferentes tipos. O microcontrolador utilizado é um ARM de 60 MHz da *Phillips* com memória *Flash* o que permite o armazenamento de dados. Existem até 16 canais com 16 bits de resolução para sinais analógicos e possibilidade de ligação de sensores digitais com interface *RS232*.

No entanto foi necessário projectar e implementar uma *housing* estanque para colocar toda a electrónica do CTD bem como um cabo que permitisse a comunicação com o exterior e ao mesmo tempo garantisse a estanquicidade do sistema. A *housing* foi feita em polipropileno enquanto que para ligação com da placa central com o PC foram usadas cabos e as fichas usadas da SUBCONN.

##### Algumas especificações Técnicas

- Tamanho: 366 x Ø 95 mm
- Alimentação: 8-30 V
- Consumo: 1,5 W / 12 V
- Comunicação: RS232 ou I<sup>2</sup>C

##### **4.1.2.1 Sensores standard**

O sistema possui 3 sensores *standard*: Temperatura, Condutividade e Pressão. O sensor de temperatura usado é uma PT100 que permite leituras desde  $-5\text{ }^{\circ}\text{C}$  a  $35\text{ }^{\circ}\text{C}$  com uma resolução de  $\pm 0,001\text{ }^{\circ}\text{C}$ . Para a aferição da Pressão é usado um sensor baseado na extensometria e tem como gama de valores de entrada 0-100 dbar e 0,2 % do fim de escala.

Para a medição da condutividade são usados eléctrodos com uma resolução de  $\pm 0,01\text{ mS/cm}$  e o seu alcance vai até os 70 mS/cm. As características estáticas dos sensores estão resumidas na tabela 6

**Tabela 6: Características dos sensores de condutividade, temperatura e pressão**

Grandeza	Condutividade	Temperatura	Pressão
Sensor	7 Electrodo	PT 10	Célula de carga
Unidades	mS/cm	° C	dbar
Alcance	0-70 mS/cm	[-5 a 35°C]	0-100 dbar
Resolução	± 0.01 mS/cm	±0,001°C	0,2 % full scale

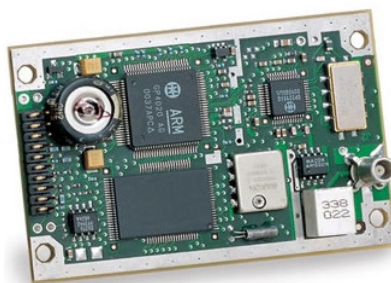
A profundidade e a salinidade são grandezas indirectas e obtidas a partir destas. A salinidade é obtida em função da condutividade enquanto que a profundidade é função da pressão.

### 4.1.3 GPS

O Sistema de Posicionamento Global (GPS) tem permitido uma grande variedade de aplicações civis e militares. O sistema é composto por uma constelação de 24 satélites distribuídos em seis planos orbitais localizados aproximadamente a 20000 km de altitude.

Os sinais enviados pelos satélites GPS são capturados na superfície terrestre por equipamentos apropriados (receptores GPS), permitindo a determinação da posição 3D; solução que só é possível se pelo menos quatro satélites forem sintonizados simultaneamente pelo receptor.

Os receptores GPS transmitem sinais em diversos formatos. Um dos protocolos mais utilizados é aquele desenvolvido e padronizado pela *National Marine Eletronic Association* (NMEA). Entre os formatos NMEA, pode-se ter várias versões, como por exemplo: NMEA 0180, NMEA 0182, e NMEA 0183.

**Figura 34 - Receptor de navegação Superstar II**

O receptor utilizado neste trabalho é um receptor de navegação da Empresa *Novatel* (*Superstar II*) e o protocolo de comunicação entre o receptor GPS e o PC obedece ao padrão

NMEA, versão 0183. As mensagens são transmitidas, em blocos (*frame*) constituídos por vários caracteres, sendo os blocos inicializados pela *string* \$GP. Como exemplo de um bloco de mensagem transmitida tem-se:

\$GPGGA,123519,4807.038,N,01131.000,E,1,08,0.9,545.4,M,46.9,M,,\*47

#### 4.1.4 Comunicação

O sistema comunica com o *PC* via porta série. Os dados são enviados com uma taxa de transmissão de 9600 bps no formato ASCII.

A comunicação do CTD é *full-duplex* e para que o sistema envie dados é necessário saber o *id* da placa. Essa informação é obtida enviando o comando “\$?<cr>”.

O CTD permite envio de dados em modo contínuo ou apenas quando solicitado através do comando. Para além disto podemos desejar receber os dados já tratados (com o cálculo da salinidade) ou “*raw data*” para posterior tratamento. Existem muitas outras funcionalidades associadas como a alteração do *id* da placa, mudança de taxa de transmissão entre outras, fazer o *download* dos coeficientes para calibração do “*raw data*” entre outras.

Em modo contínuo os dados são enviados em intervalos de tempo de 100 ms e possuem um formato próprio.

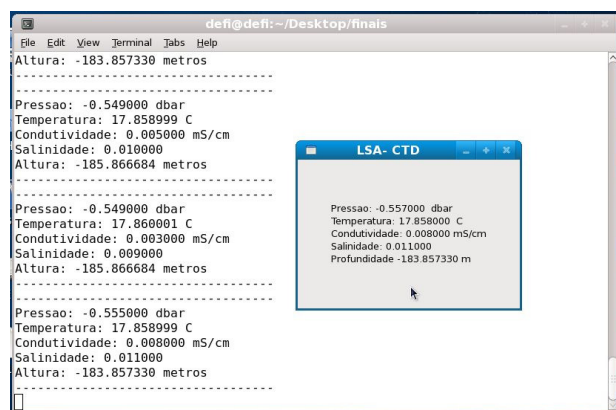
!id#+ppp.ppp,+tt.ttt,+cc.ccc,+ss.ss<cr>

- *id* (2-3)- *Id* da placa
- *ppp.ppp* (5-12): Valor da pressão absoluta (dbar)
- *tt.ttt* (14-20): Valor da temperatura (°C)
- *cc.ccc* (22-28): Valor da condutividade (mS/cm)
- *ss.sss* (30-35): Valor da salinidade
- <cr>: *Carriage Return* (0D em hexadecimal)

O GPS envia continuamente uma série de tramas, com múltiplos parâmetros a uma taxa de 5 Hz. No entanto apenas uma nos interessa e que contém a posição da latitude, longitude e altura e que tem início com a *string* \$GPGGA

\$GPGGA,123519,4807.038,N,01131.000,E,1,08,0.9,545.4,M,46.9,M,,\*47

Foi usado o sistema operativo baseado em *Linux Fedora 10* e a aquisição de dados feita em Linguagem *C* tendo sido desenvolvida uma interface gráfica em *GTK+* (figura 35) para fazer a monitorização em tempo real dos dados recolhidos. Os dados recolhidos são guardados num *log* com a indicação do valor das diferentes grandezas bem como a posição (latitude, longitude e altura) onde foi obtida.



**Figura 35 - Interface gráfica com o valor das grandezas**

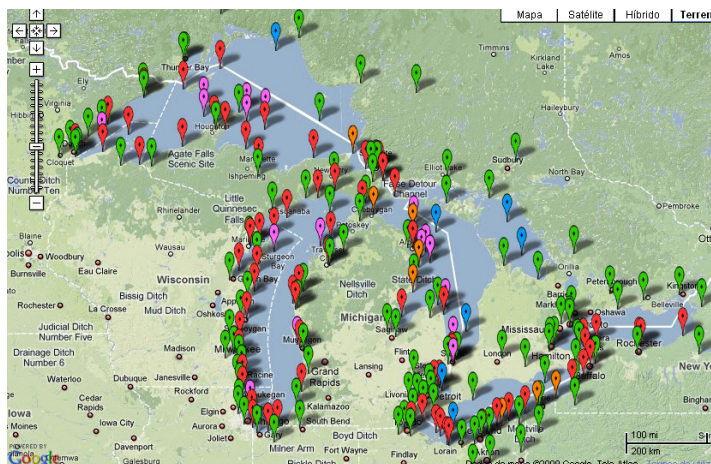
*GTK* é uma abreviação para *GIMP Tool Kit*. O *GIMP* (*GNU Image Manipulation Program*) é um editor gráfico, com licença *GPL*, o que significa que é gratuito, e pode ser distribuído livremente e foi concebido em *C*, para ser usado na plataforma *Linux*, sendo fácil a sua utilização.

## 4.2 Amostragem

O protótipo foi desenvolvido e encontra-se pronto para se tornar operacional, no entanto não foi possível efectuar qualquer missão, em tempo útil, devido à instabilidade climática. Devido a esse facto a amostragem dos dados foi feita recorrendo ao projecto *Costwatch* da *NOOA*.

O *NOAA* (*National Oceanic & Atmospheric Administration*) é uma organização estadual americana cuja missão é a de monitorização atmosférica e oceânica, com vista à salvaguarda de vidas e bens. O *Coastwatch* é um programa de monitorização ambiental dos Grandes Lagos do *GLERL* (*Great Lakes Environmental Research Laboratory*) e do *NOAA* que produz e fornece dados ambientais para o apoio à ciência ambiental através de observações via satélite e dados *in-situ* na região dos Grandes Lagos (Figura 36).

Os dados *in situ* são recolhidos através de bóias, estações meteorológicas costeiras e incluem observações como a velocidade e direcção do vento, temperatura do ar de ponto de orvalho, cobertura de nuvens, a pressão do nível do mar, temperatura da água, altura de onda e período de onda.



**Figura 36 - Localização das estações do Programa Costwatch**

### 4.3 Metodologia

Os dados utilizados para o modelo geoestatístico foram obtidos através do portal do Programa Costwatch (<http://coastwatch.glerl.noaa.gov/marobs/marobs.html>) e foram recolhidos um conjunto de dados de temperatura relativos ao primeiro dia do mês de Agosto de 2009 com uma separação de 4 horas.

A análise tem como objectivo a estimação do valor da temperatura para um dos lagos (*Lake Erie*), para duas situações distintas utilizando as estações disponíveis na área de estudo (Figura 37). Na primeira recorre-se à estimação do valor de temperatura recorrendo a valores recolhidos em nove bóias ao longo do lago. Para provar as vantagens do uso de veículos móveis utiliza-se os valores recolhidos nas estações costeiras, que se irá utilizar para simular o trajecto de um veículo, para estimação do valor de temperatura.

Em ambas as situações são utilizados três pontos (Sensores Virtuais) distintos cujo valor é conhecido e comparado com o valor estimado para ambas as situações.



**Figura 37 - Localização das estações ao longo do *Lake Erie***

## 4.4 Desenvolvimento

Para a manipulação e análise de dados e para a geração de mapas foi usado o software *MATLAB* que é um pacote comercial para a análise numérica, desenvolvido por *The Mathworks, Inc.* O *MATLAB* possui uma série de comandos versáteis para operações matemáticas, estatísticas e produção de gráficos de alta qualidade. O *MATLAB* possui uma série de *toolboxes*, que permitem a criação de funções e programas reutilizáveis, tornando a ferramenta versátil.

Os dados recolhidos de cada estação são apresentados num formato próprio, dando ao utilizador informação temporal (ano,dia,hora,minuto), do tipo de estação (bóia, estação costeira, etc...), o identificador da estação e finalmente o valor dos diversos parâmetros físicos monitorizados.

A fase seguinte foi a recolha da posição georreferenciada de cada estação que se apresenta com dois sistemas de coordenadas (Graus – Minutos - Segundos e Graus decimais). No sistema Graus - Minutos - Segundos , cada grau é dividido em 60 minutos, que por sua vez se subdividem, cada um, em 60 segundos ( $47^{\circ} 49.2' N$   $83^{\circ} 21.0' W$  ). No sistema Graus Decimais, cada grau é dividido em fracções decimais existindo valores positivos para estações a Norte (latitude) e a Este (longitude) e negativos para localizações a Sul (latitude) e a Oeste (longitude).

Os dados recolhidos foram colocados em ficheiros (extensão *.txt*) com o seguinte formato:

```
Id,+lat,+lng,temp <cr>
```



- *id*- Identificador da estação
- *lat*- Latitude (Graus decimais)
- *lng*- longitude (Graus decimais)
- *temp*- Valor da temperatura(°C)

No entanto para saber distâncias entre estações é necessário ainda a conversão do sistema de coordenadas elipsoidais (latitude, longitude) usado pelo programa *Costwatch* para o sistema cartesiano (X,Y,Z). Foi desenvolvida a função `conv_gps.m` que recebe os valores da localização da estação em coordenadas elipsoidais e as converte em cartesianas.

No entanto esta posição dá-nos ainda a distância euclidiana (em km) em relação centro da terra. É por isso necessário a indicação de um ponto de referência para determinar a localização de cada estação. A função `posicao_xyz` faz essa conversão, tornando mais fácil o tratamento e manipulação de dados.

```
[coord, z] = posicao_xyz (fich_xyz, lat_ref, lng_ref, h_ref)
```

- *fich\_xyz*- Ficheiro com a localização do ponto (Graus decimais) e respectivo valor da grandeza;
- *lat\_ref*- Valor da latitude do ponto de referência;
- *lng\_ref*- Valor da longitude do ponto de referência;
- *h\_ref*- Valor da altura do ponto de referência.

Parâmetros de saída

- *coord*- Valor das coordenadas em relação ao ponto de referência;
- *z*- Valor da grandeza monitorizada.

Para este trabalho foram desenvolvidas uma série de rotinas e funções baseadas na *toolbox* de análise espacio-temporal *BMELib* (*Bayesian Maximum Entropy Library*) [39].

Esta *toolbox* utiliza ficheiros com um formato de texto *standard* (padrão *GeoEAS* da *US-EPA*) conforme apresentado de seguida:

```
Início do ficheiro -----
Temperatura no ISEP às 00:00
3
x (Km) , x-coordenadas em Km
y (Km), y coordenadas em Km
Temperatura
10.1    11.4    10.2
12.2    13.6    12.2
16.7    19.1    12.4
10.9    16.9    11.8
Fim do ficheiro -----
```

Para colocar os dados em ficheiros (extensão *.dat*) neste formato utilizou-se a rotina `writeGeoEAS.m` e para a leitura destes dados a `readGEOEAS.m`.

A primeira etapa na modelização é a análise exploratória de dados, onde é feita uma análise estatística descritiva dos dados obtidos. A função `analise_dados.m` faz a leitura dos dados (Padrão *GeoEAS*) e dá-nos a indicação da média, variância, desvio padrão, valores mínimos e máximos da grandeza medida. Esta rotina gera ainda um mapa com a localização espacial dos pontos recolhidos e o respectivo valor da grandeza bem como histograma dos pontos. Essa informação é colocada num ficheiro extensão XML para posterior disponibilização de dados via *Web*.

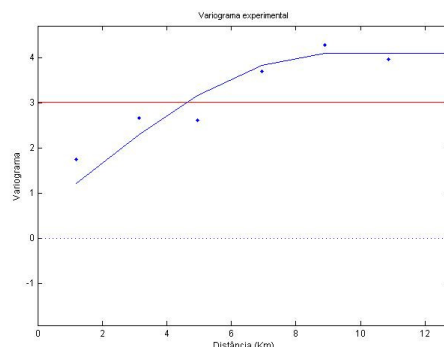
```
<markers>
<marker lat="42.27" lng="-82.97" valor="24.0" />
<marker lat="41.780" lng="-80.70" valor="19.0" />
<marker lat="43.03" lng="-81.15" valor="20.0"/>
</markers>
```

**Figura 38: Exemplo do ficheiro XML com os pontos amostrados e respectivos valores da temperatura.**

O XML é um formato para a criação de documentos com dados organizados de forma hierárquica que não depende das plataformas de *hardware* ou de *software* tornando-o portátil. Assim uma base de dados pode através de uma aplicação, escrever num ficheiro XML, e uma outra pode ler esses mesmos dados.

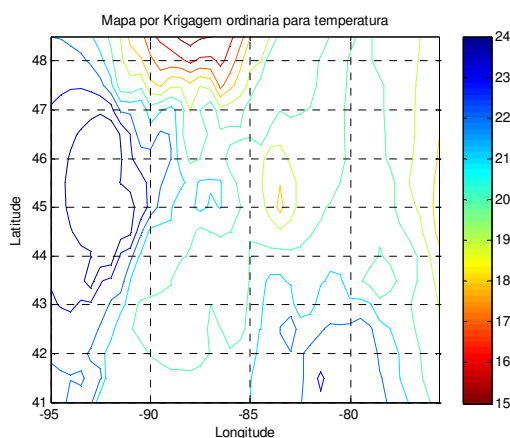
O semivariograma, ferramenta básica de suporte às técnicas de geoestatística que permite representar quantitativamente a variação de um fenómeno regionalizado foi implementada através da rotina `semivar.m`. Esta lê o *log* de dados e recebe como parâmetros de entrada o número de pontos que quero usar para o cálculo, o tipo de *krigagem* a usar, o modelo que quero usar bem como os ângulos iniciais e finais para verificar se existe alguma tendência. Os valores do efeito pepita, patamar e alcance são também usados para traçar a

curva do semivariograma. Os pontos aferidos para o semivariograma são igualmente guardados em formato XML

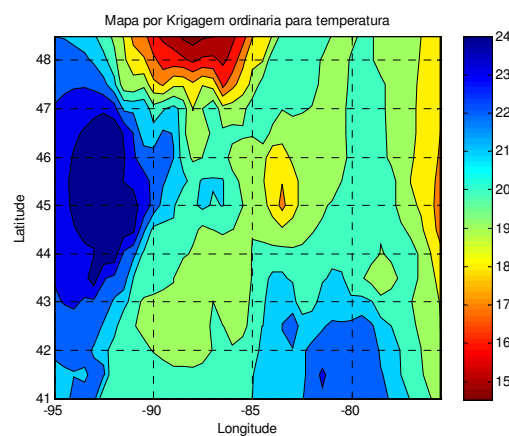


**Figura 39- Exemplo de um variograma experimental**

Finalmente e para a geração de mapas implementou-se a rotina `krig.m`. Para tal deve-se definir a malha de interpolação, limites máximos e mínimos dessa malha e o espaçamento de pontos ou número de linhas. Logicamente uma malha mais densa terá um tempo de processamento maior.

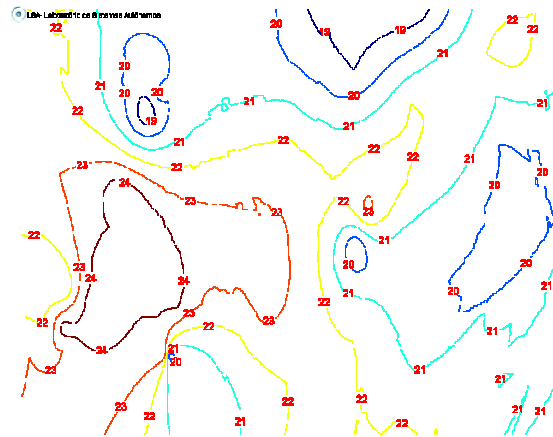


**Figura 40: Exemplo de mapa gerado**



**Figura 41 – Outro exemplo de um mapa gerado**

A pensar na utilização via *Web* destes mapas recorrendo a plataforma do *Google Maps* fez-se a adaptação da rotina anterior com vista a geração de mapas com fundo transparente conforme se pode ver na figura 42.



**Figura 42: Mapa gerado com valores de isoterma.**

Para efectuar testes de verificação usa-se três pontos de teste aleatórios, mas cujo valor é conhecido e efectua-se a simulação para diferentes situações. A rotina implementada retorna o valor dos erros esses 3 pontos para cada uma das simulações

$[e1\ e2\ e3\ e4\ e5\ e6] = \text{mapas\_erros}(\text{fich\_boia}, \text{fich\_veiculo}, \text{fich\_pontos})$

- *fich\_boias*- Ficheiro com os pontos de temperatura relativos às estações fixas;
- *fich\_veiculo*- Ficheiro com os pontos de temperatura recolhidos pelas bóias e pelo veículo autónomo;
- *fich\_pontos*- Ficheiro com pontos de temperatura de estações fixas que servirão como pontos de teste.

A função retorna os erros de cada simulação.

- *e1, e2 e3*- Erros relativos aos três pontos usando os pontos do *fich\_boia*;
- *e4, e5 e6*- Erros relativos aos três pontos usando os pontos do *fich\_veiculo*.

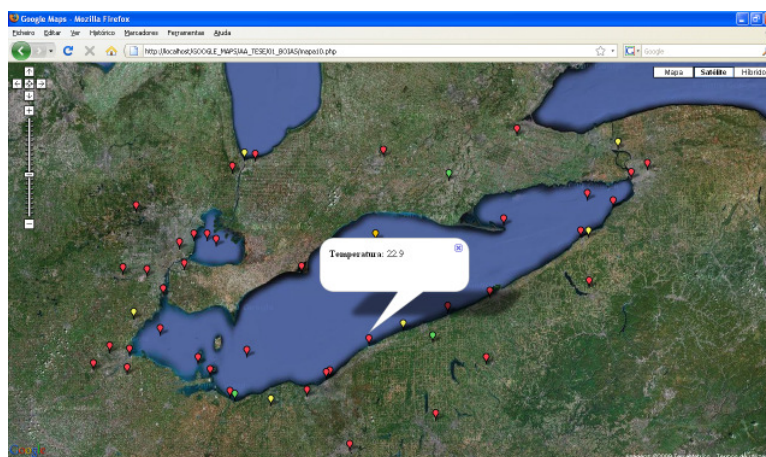
#### 4.4.1 Plataforma Web

Actualmente a internet é um veículo popular para distribuição de informação, apresentando-se como um poderoso ambiente para partilhar dados.

Com o objectivo de visualizar os dados georreferenciados em ambiente *Web* provenientes das estações e dos sistemas móveis foi desenvolvida uma aplicação baseada no *Google Maps* e *Google Earth*.

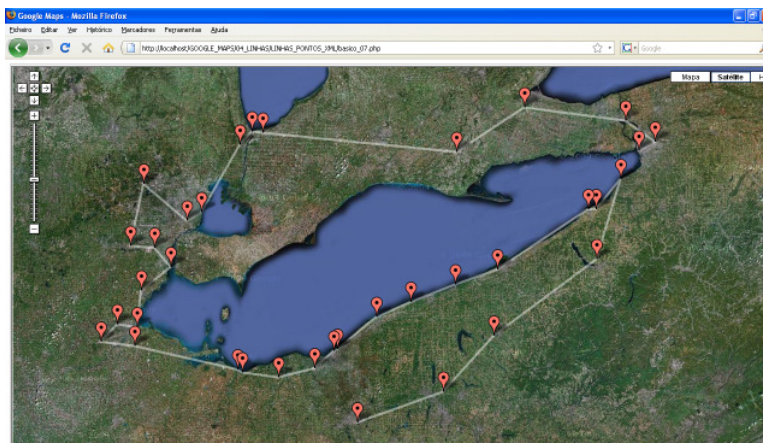
O *Google Maps* é um serviço de geoprocessamento que usa uma poderosa tecnologia de mapas e com uma interface amigável. A API do *Google Maps* é uma interface de desenvolvimento que utiliza *Javascript* para incorporar os seus mapas numa página da *Web* necessitando no entanto de uma chave que só é válida para um directório ou domínio. A API fornece diversas funções para manipular mapas e adicionar conteúdos ao mapa através de diversos serviços. A representação dos dados é feita em linguagem XML que oferece uma abordagem padrão para descrição, captura, processamento e publicação de informações

A aplicação foi programada em linguagem HTML, PHP e JAVA, utilizando a API do *Google Maps* e XML para a transferência de informação. (figura 43). Os dados depois de recolhidos são tratados no MATLAB e gerado um ficheiro XML com a localização das estações e respectiva temperatura.



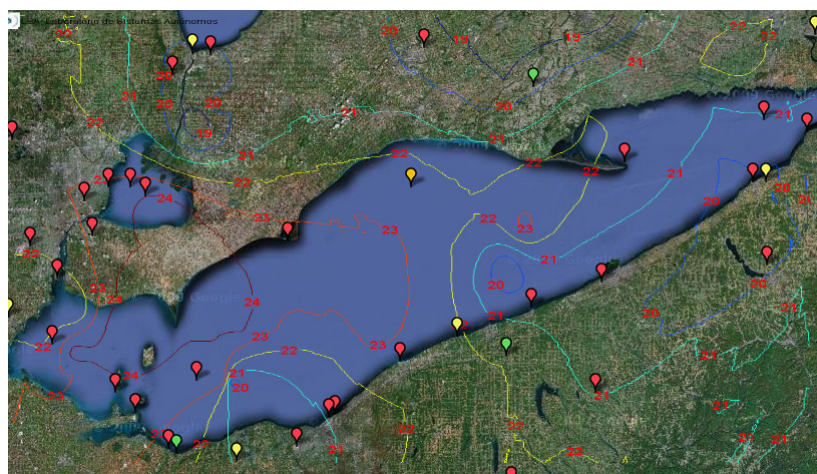
**Figura 43 - Representação dos pontos recolhidos por estações fixas e veículos móveis**

Com esta ferramenta é no entanto possível visualizar, em simultâneo, o trajecto efectuado por qualquer veículo autónomo como se pode ver na figura 44.



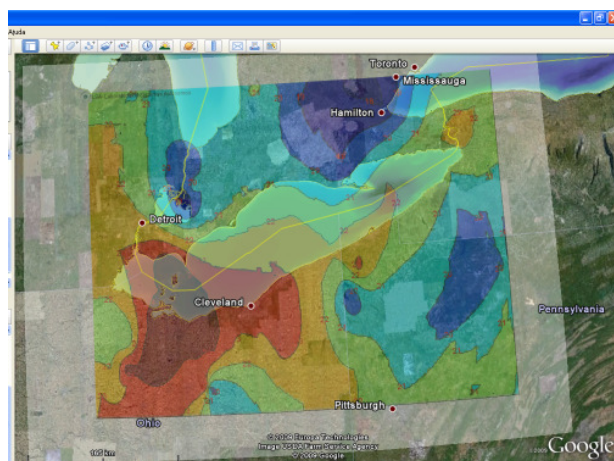
**Figura 44 - Representação dos pontos recolhidos pelo veículo móvel e respectivo percurso.**

Depois do tratamento dos dados e gerado o mapa pelo *MATLAB* é sobreposta recorrendo a um conjunto de funções



**Figura 45 - Representação dos pontos recolhidos pelo veículo móvel e respectivo percurso.**

Outra das ferramentas disponíveis é o *Google Earth* que não é mais do que um programa desenvolvido e distribuído pelo *Google* cuja função é apresentar um modelo tridimensional do globo terrestre. (Figura 46)



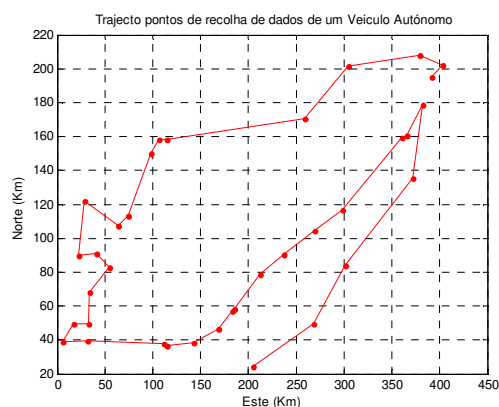
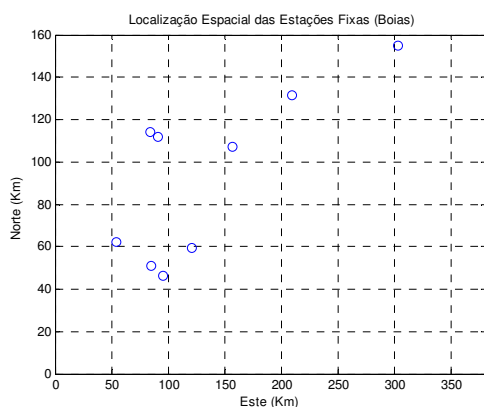
**Figura 46 - Aplicação *Google Earth* com a sobreposição do mapa gerado.**

**Esta página foi intencionalmente deixada em branco**



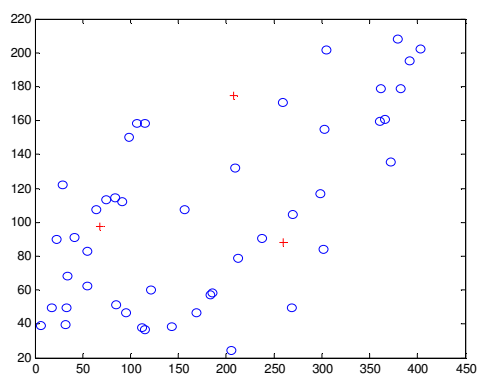
## 5 Resultados Obtidos

Na figura 47 podemos ver a localização espacial das bóias utilizadas para simulação de uma rede de sensores fixos constituída por nove estações, bem como o trajecto do veículo autónomo e respectivas amostras recolhidas (figura 48).



**Figura 47 - Localização espacial das estações fixas. Figura 48 – Trajecto efectuado por um veículo móvel.**

Os pontos escolhidos como pontos de teste estão assinalados (+) na figura 49 e são três estações que serviram para determinação dos parâmetros do modelo, mas que não farão parte de qualquer modelo estimado.



**Figura 49 - Localização espacial de todos os pontos recolhidos.**

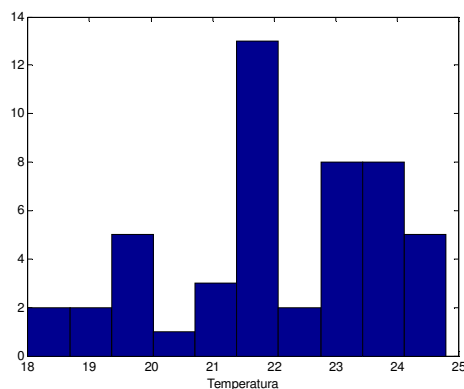
## 5.1 Análise exploratória de dados

A primeira fase da modelização geoestatística é a análise exploratória de dados ou seja a avaliação do comportamento dos dados disponíveis. A tabela 7 apresenta o resumo estatístico do conjunto de dados de temperaturas (50 medidas) recolhidas a 1 de Agosto de 2009 às 00:00 Horas.

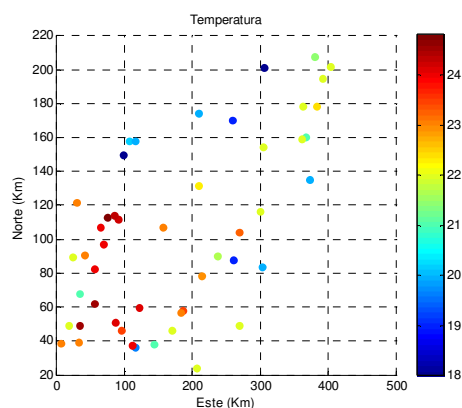
**Tabela 7: Resumo estatístico dos dados das medidas de temperatura**

Parâmetro	Temperatura ( ° C)
Mínimo	18.0
Média	22.1327
Máximo	24.8
Variância	3.0322
Desvio Padrão	1.7413

O valor da temperatura varia entre 18.0 °C e 24.8 °C, sendo o valor médio de 22.1327 °C. A figura 50 mostra o histograma do conjunto de dados da temperatura, enquanto que na figura 51 podemos ver a respectiva distribuição espacial dos dados



**Figura 50 – Histograma de pontos referentes à temperatura no dia 01 de Agosto de 2009 (00:00)**



**Figura 51 – Respectiva distribuição espacial**

## 5.2 Parametrização

A figura 52 mostra ao semivariograma experimental omnidireccional da temperatura ajustado ao modelo esférico.

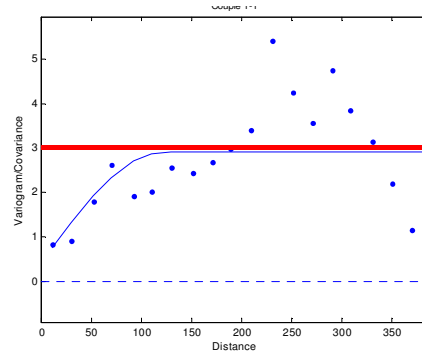


Figura 52 –Semivariograma experimental Omnidireccional da temperatura

Definido por

$$\gamma(\mathbf{h}) = \begin{cases} 3 \left( 1.5 \frac{400}{100} - 0.5 \frac{400^3}{100^3} \right) & h \leq 100 \\ 3 & h > 100 \end{cases} \quad (5.1)$$

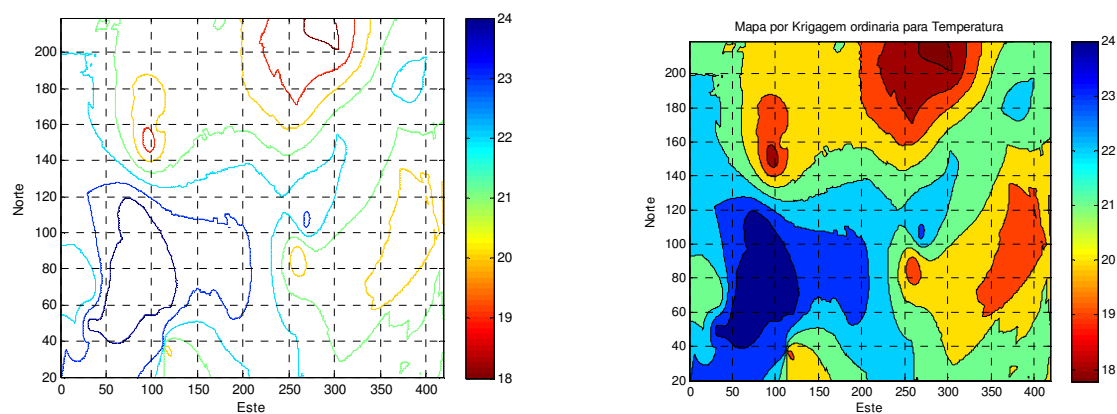
Os parâmetros *pepita*, *patamar* e *alcance* do modelo ajustado para o semivariograma da temperatura são mostrados na tabela 8.

Tabela 8 – Parâmetros do modelo ajustado para o semivariograma da temperatura

Modelo	Esférico
Pepita	0
Patamar	3
Alcance	100

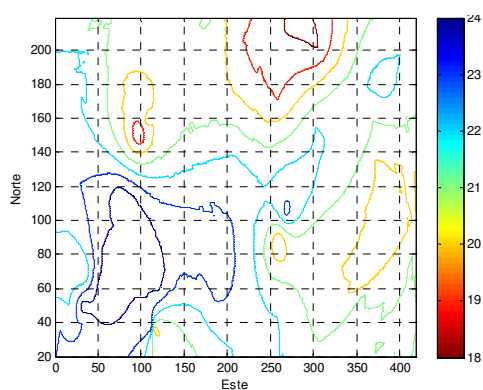
## 5.3 Estimação e Geração de Mapas

O mapa da distribuição da temperatura na zona dos Grandes Lagos, mais precisamente no *Lake Erie*, utilizando o modelo de semivariograma esférico, pode-se ver na figura 54 e 55

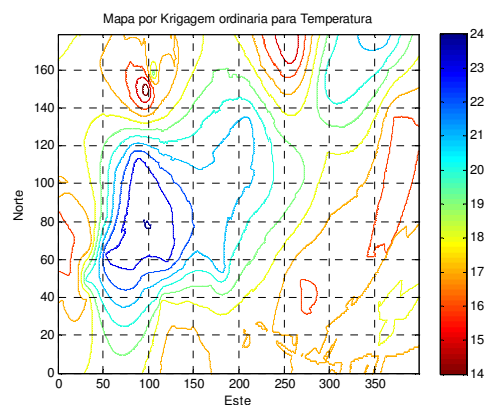


**Figura 53 - Mapa da distribuição da temperatura na zona dos Grandes Lagos (*Lake Erie*).**

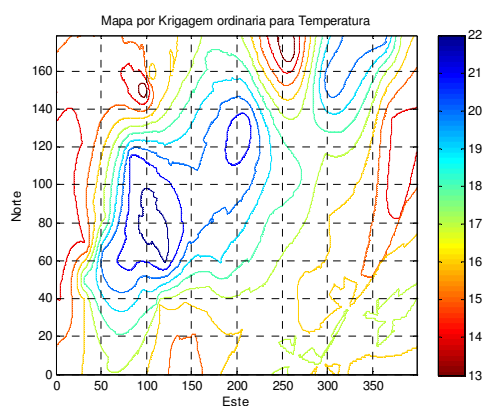
São apresentados aqui os mapas da evolução da temperatura ao longo do dia utilizando pontos fixos com ajuda da informação proveniente de veículos móveis (EF+VM) para o dia 01 de Agosto de 2009.



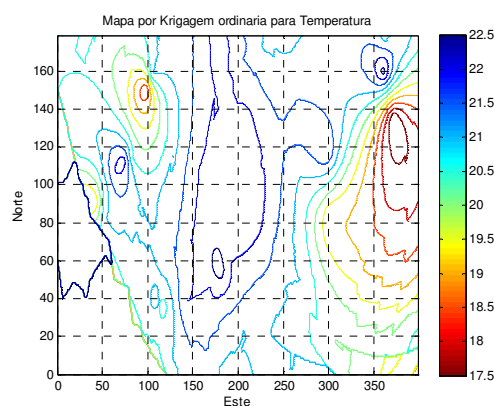
**Figura 54- Mapa da distribuição da temperatura às 00:00**



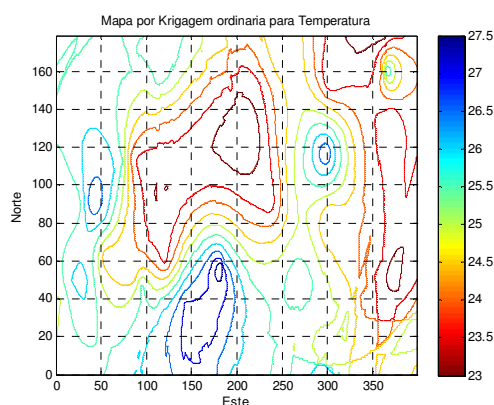
**Figura 55 - Mapa da distribuição da temperatura às 04:00**



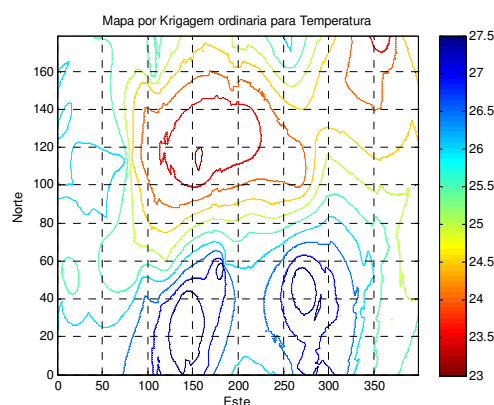
**Figura 56 - Mapa da distribuição da temperatura às 08:00**



**Figura 57 - Mapa da distribuição da temperatura às 12:00**



**Figura 58 - Mapa da distribuição da temperatura às 16:00**



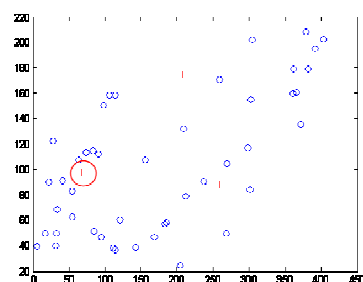
**Figura 59 - Mapa da distribuição da temperatura às 20:00**

## 5.4 Avaliação de Erros

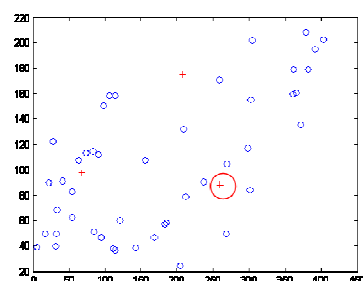
Apresentam-se aqui os erros relativos aos pontos de teste estimados quer com dados provenientes de sensores fixos (EF) quer usando dados de todos os pontos disponíveis ou seja provenientes de sensores fixos e móveis (EF+VM), para diferentes horas do dia 01 de Agosto de 2009.

**Tabela 9- Erros relativos ao ponto 1**

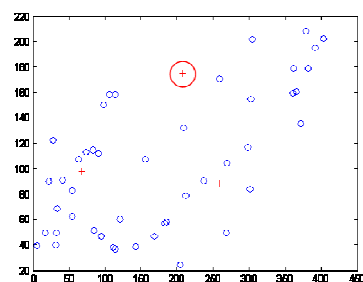
Ponto	Valor (°C)	Erro (EF)	Erro (EF+VM)
00:00	24.0	0.3755	0.4918
04:00	20.0	3.0017	1.7005
08:00	17.0	4.040	1.9758
12:00	21.0	0.5575	0.5560
16:00	26.0	2.1374	0.4645
20:00	26.0	1.4287	0.2514

**Tabela 10- Erros relativos ao ponto 2**

Ponto	Valor (°C)	Erro (EF)	Erro (EF+VM)
00:00	19.0	3.5257	2.5058
04:00	16.0	5.6342	2.7984
08:00	16.0	5.5010	1.7345
12:00	21.0	0.5027	0.1249
16:00	26.0	3.1582	1.3827
20:00	27.0	3.3156	2.0518

**Tabela 11- Erros relativos ao ponto 3**

Ponto	Valor (°C)	Erro (EF)	Erro (EF+VM)
00:00	20.0	2.6613	0.9727
04:00	15.0	6.6810	2.7886
08:00	14.0	7.6710	3.2095
12:00	21.0	0.7078	0.5555
16:00	24.0	1.1971	0.0362
20:00	24.0	0.4370	1.0717



## 6 Conclusões e Trabalhos Futuros

Nesta dissertação apresentamos uma solução para um veículo autónomo aquático de superfície. Esta permite a recolha de condutividade, temperatura, salinidade e profundidade quer para ambientes marítimos quer fluviais. Esta solução integra ainda um receptor GPS para que os dados recolhidos sejam georreferenciados e guardados num *log* de dados.

As sucessivas etapas que conduziram à solução final integrada (figura 31) foram:

- Estudo do funcionamento dos sensores de condutividade, temperatura e pressão;
- Desenvolvimento do *housing* que permita integrar toda a electrónica a garantindo a estanquicidade à água;
- Testes de estanquicidade efectuados nas instalações do LSA, até 5 m de profundidade;
- Estudo do protocolo de comunicação do CTD e do GPS;
- Desenvolvimento de software em Linguagem C para a aquisição de dados e usando como sistema operativo o *Linux Fedora 10*;
- Desenvolvimento de uma interface gráfica em *GTK+* para a monitorização em tempo real dos dados recolhidos.

O protótipo desenvolvido tem as seguintes características

- Tamanho: 366 x Ø 95 mm
- Alimentação: 12 V
- Consumo: 1,5 W / 12 V
- Comunicação: RS232

A solução apresentada satisfaz os requisitos, pois permite a realização de missões de observação de fenómenos em rios, lagos e oceanos, de uma forma ligeira e a custos reduzidos, permitindo o registo de dados georreferenciados

Para esta tese foi desenvolvido um conjunto de funções em *MATLAB* que implementam os algoritmos dos modelos geoestatísticos. As rotinas implementadas serviram para fazer a análise estatística dos dados, para a determinação dos parâmetros a usar no modelo e para estimação de valores e consequente geração dos mapas.

Os dados utilizados neste trabalho foram obtidos através do programa *CoastWatch* (Figura 36) e permitiram a simulação da utilização de sistemas móveis, em complemento a sistemas fixos.

Foi definida a estratégia de observação que permitisse obter uma informação caracterizadora do meio através da trajectória para o veículo móvel (Figura 48), sabendo à *priori* a localização das estações fixas.

Com os dados obtidos foram calculados os indicadores estatísticos clássicos do conjunto de observações de temperatura e depois do cálculo do semivariograma experimental chegou-se à conclusão que não foi detectada nenhuma direcção específica, mostrando-se o fenómeno ser isotrópico.

Foram então utilizados os parâmetros indicados na tabela 8 no modelo ajustado para o semivariograma da temperatura. A Figura 53 mostra o resultado das previsões de temperatura, obtidas usando o modelo esférico e utilizando a *krigagem* ordinária.

Para avaliar a abordagem metodológica, ou seja como é que os modelos geoestatísticos, com calibrações frequentes com veículos móveis, podem ou não apresentar melhores resultados, foram recolhidos dados do dia 1 de Agosto de 2009 com intervalos temporais de 4 horas. Foram ainda recolhidos valores de três pontos que serviram de teste para quantificar o erro proveniente das simulações.

Com estes dados foram elaborados os mapas de distribuição de temperatura utilizando apenas os sensores fixos e uma outra usando adicionalmente os valores recolhidos pelo veículo móvel (Figuras 54 a 59) para os diferentes intervalos de tempo.

Os erros obtidos na estimação dos três pontos que serviram de teste, mas que não entraram na simulação de qualquer caso são mostrados nas tabelas 9 a 11.

Como se pode constatar o erro obtido para cada ponto, a partir do modelo geoestatístico, é muito menor quando possui informações de ambos os sensores (fixos e móveis).

Os resultados preliminares obtidos sustentam a convicção do uso de veículos móveis na estimação de grandezas em tempo-real, pois desta forma diminuirá o erro obtido na estimação de qualquer ponto da área em estudo, permitindo uma melhor caracterização dos fenómenos naturais.

A análise dos resultados do trabalho apresentado sugere claramente duas formas de dar continuidade aos desenvolvimentos já efectuados. A primeira através da validação desta metodologia com dados provenientes da solução para um veículo autónomo de superfície apresentada no capítulo 4. A segunda na introdução de outras variáveis, como relevo, direcção



e velocidade do vento, na estimação da grandeza e a respectiva validação metodológica proposta nesta tese.

Outro aspecto a ter em conta em trabalhos futuros é a forma como os resultados são afectados pelo facto da rede possuir sensores com diferentes precisões e resoluções.

Perspectiva-se ainda, num futuro, a utilização da geoestatística na concepção e planeamento de missões de veículos móveis, tendo o prévio conhecimento a localização dos sensores que a rede possui, isto para evitar um elevado número de dados numa área em detrimento de outras que não possuam qualquer cobertura.

Esta página foi intencionalmente deixada em branco.

## 7 Bibliografia

- [1] Haenggi,, “Opportunities and Challenges in Wireless Sensor Network,” *Handbook of Sensor Networks Compact wireless and Wired Sensing Systems*, CRC press, 2005.
- [2] Mainwaring, Alan, Polastre, Joseph, Culler, David, e Anderson, John, “Wireless Sensor Networks for Habitat Monitoring,” *WSNA-ACM*, Out. 2002.
- [3] Holzmman, Michael e Kirsch, Christoph, “Wireless Sensor Networks and Their Application A Survey,” *Seminar Computational Systems SS 2004*, Jul. 2004.
- [4] R.L. Beatrys, L. Correia, L. Vieira, D. Macedo, E. Nakamura, C. Figueiredo, M. Vieira, Eduardo Mechelane, D. Camara, A. Loureiro, J. Nogueira, e D. Jr., “Arquitecturas para redes de sensores sem fio,” *Livro texto da XXI Jornada de Atualização em Informática do XXII Congresso da Sociedade Brasileira de Computação*, 2004.
- [5] Silva, Fabrício Aguiar, Braga, Thais Regina de Moura, Ruiz, Linnyer Beatrys, e Nogueira, José Marcos Silva, “Tecnologia de Nós Sensores Sem Fio.”
- [6] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, e D. Estrin, “Habitat Monitoring With Sensor Networks,” *Communications Of The ACM*, vol. 6, 2004.
- [7] A. Baggio, “Wireless Sensor Networks in Precision Agriculture.”
- [8] J.M.S. Nogueira, Ruiz, Linnyer Beatrys, e R.A.D.F. Mini, “Redes de Sensores Sem Fio- Capitulo 5 ,” *Livro texto da XXI Jornada de Atualização em Informática do XXII Congresso da Sociedade Brasileira de Computação*, 2004.
- [9] M.D.C. Barreto e E.B. Ornellas, “Redes de Sensores Sem Fio: Análise de Sua Implementação no Monitoramento Ambiental,” Mai. 2007.
- [10] M.R. Pereira, C.L.D. Amorim, e M.C.S.D. Castro, “Tutorial sobre Redes de Sensores.”
- [11] Ruiz, Linnyer Beatrys, “MANÁ: Uma Arquitetura para Gerenciamento de Redes de Sensores Sem Fio,” *Tese Pós-graduação, Universidade Federal de Minas Gerais*, 2003.
- [12] S.A.F.D. Silva e J.R.M.D. Silva, “Agricultura de Precisão – Análise Espacial Comparativa da Produtividade do Milho, para Diferentes Rampas Rotativas na Herdade do Cego,”
- [13] John Orcutt e John Helly, “Sensor Networks - The Design and Implementation of Aquatic and Marine Sensor Marine Sensor Networks,” *Sine Workshop Report*.
- [14] Ian F. Akyildiz e Ismail H. Kasimoglu, “Wireless Sensor and Actor Networks: Research Challenges,” *Ad Hoc Networks 2*, 2004, pp. 351–367.
- [15] Sameer Tilak, Nael B. Abu-Ghazaleh, e Wendi Heinzelman, “A Taxonomy of Wireless Micro-sensor Network Models,” *Mobile Computing and Communications Review*, v1
- [16] S. Megerian, F. Koushanfar, Gang Qu, Giacomino Veltri, e Miodrag Potkonjak, “Exposure In Wireless Sensor Networks: Theory And Practical Solutions,” 2002.

- [17] Rodrigo Portella do Canto, *SensorNet – Uma Proposta de Aplicação para Gerenciamento de Redes de Sensores Sem Fio*, Universidade Federal do Rio Grande do Sul, 2008.
- [18] A. SOARES, *Geoestatística para as ciências da terra e do ambiente*, Ist Press - Instituto Superior Técnico , 1999.
- [19] M.P. Nascimento, “Redes de Sensores,” 2003.
- [20] G. Sibley, M. Rahimi, e G. Sukhatme, “Robomote: A Tiny Mobile Robot Platform for Large- Scale Sensor Networks,” *IEEE Int’l Conf. Robotics and Automation*, 2002.
- [21] L.F.M. Vieira, “Middleware para Sistemas Embutidos e Redes de Sensores,” Tese de Mestrado, Belo Horizonte, 2004.
- [22] A. LaMarca, W. Brunette, D. Koizumi, M. Lease, S. Sigurdsson, K. Sikorski, D. Fox, e G. Boriello, “Making sensor networks practical with robots,” *International Conference on Pervasive Computing*, 2002.
- [23] A. Boulis, “Models for Programmability in Sensor Networks,” *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, CRC Press, 2005.
- [24] P. Stein, “Practical Considerations for Environmental Monitoring with Wireless Sensor Networks,” Mai. 2004.
- [25] J.K. Hart e K. Martinez, “Environmental Sensor Networks: A revolution in the earth system science?,” *Elsevie*, 2006.
- [26] J. Ryan, F. Chavez, J. Bellingham, E. Rienecker, H. Dierssen, R. Kudela, A.V. Woude, R. Maffione, e A. Fischer, “Environmental Processes in the Monterey Bay National Marine Sanctuary: Studies Integrating AVIRIS and Synoptic In Situ Sensing,” *Proceedings of the 2004 JPL Airborne Earth Science Workshop*, 2004.
- [27] T. Melodia, D. Pompili, V.C. Gungor, e I.F. Akyildiz, “A distributed coordination framework for wireless sensor and actor networks,” *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, 2005.
- [28] B. Argrow, D. Lawrence, e E. Rasmussen, “Uav systems for sensor dispersal, telemetry, and visualization in hazardous environments,” *43rd Aerospace Sciences Meeting and Exhibit*, 2005.
- [29] M. Rahimi, R. Pon, W.J. Kaiser, G.S. Sukhatme, D. Estrin, e M. Srivastava, “Adaptive Sampling for Environmental Robotics,” *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International* , 2004.
- [30] M.A. Ferreira, “Sistema de Monitoramento Contínuo de Qualidade de Águas Produtivas e de Mananciais,” Doutorado, São Paulo, 2003.
- [31] R. Jedermann e Lang, “The Minimum Number of Sensors – Interpolation of Spatial Temperature Profiles in Chilled Transports,” *Proceedings of the 6th European Conference on Wireless Sensor Networks*, 2009.
- [32] C. Santos, J. Uva, e J. Almeida, “Caracterização espacial da produtividade do pinheiro-bravo em Portugal,” Tagus Park, Oeiras: 2002.

- [33] B. Deus, P. Zeilhofer, e E.S. Santos, “Interpolação Espacial de Precipitações Anuais na Bacia do Alto e Médio Teles Pires,” Florianópolis: 2007.
- [34] J. Borssoi, M. Uribe-Opazo, e M.G. Rojas, “Influência Local em Análise Geoestatística,” 2008.
- [35] B. Deus, P. Zeilhofer, e E.S. Santos, “Interpolação Espacial de Precipitações Anuais na Bacia do Alto e Médio Teles Pires,” Florianópolis: 2007.
- [36] Burgess, T.M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. The Journal Soil Science, Oxford, 31:315-331, 1980.
- [37] Métodos geoestatísticos para a avaliação e gestão de recursos Minerais, Departamento de Minas IST, 1992
- [38] Jorge Kazuo Yamamoto, “Avaliação e classificação de reservas minerais”, Volume 38 da Academia de São Paulo Brasil, Editora EdUSP, 2001
- [39] Christakos, Bogaert, Serre, 2002  
*<http://www.unc.edu/depts/case/BMELIB/>*
- [40] Landim P.M.B. 2006. Sobre Geoestatística e mapas. Revista Terræ Didática, 2:19-33.  
*<http://www.ige.unicamp.br/terradidatica/>*
- [41] PROJECTO ALERT, *<http://www.alertsystems.org/>*
- [42] Vieira, Sidney, Neto. Francisco, “Variabilidade espacial do potencial de erosão das chuvas do estado de São Paulo” Bragantia, Campinas, 1995
- [43] Instituto de Oceanografia, *<http://www.io.fc.ul.pt/fisica/equipamento.htm>*
- [44] “Geoestatística Básica e Aplicada”, Ednaldo Carvalho Guimarães, Fevereiro – 2004, Uberlândia – MG
- [45] “Introdução ao Geoprocessamento”, Capítulo 8- SER 300 ao Geoestatística para geoprocessamento, Instituto Nacional de Pesquisas Espaciais (INPE), Brasil
- [46] Projecto CORIE , *<http://www.stccmop.org/CORIE/>*
- [47] Camargo E., Druck S., Câmara G, “Análise espacial de dados geográficos” Cap. 3 Análise Espacial de Superfícies. (Camargo E., Druck S., Câmara G, 2004)
- [48] Projecto PODS, *<http://www.botany.hawaii.edu/pods/overview.htm>*
- [49] Oceanus atlanticus, *<http://oceanusatlanticus.blogspot.com/2008/04/o-mar.html>*
- [50] Clark. Isobel, “Practical Geostatistics”
- [51] Ladim. Paulo, Monteiro. Rubens, “Introdução à confecção de mapas pelo software SURFER®”, UNESP/campus de Rio Claro, 2002

- [52] Camargo E., Druck S., Câmara G, “Análise espacial de dados geográficos” -Análise Espacial de Superfícies, 2004
- [53] Ricardo Braga, “Agricultura de Precisão, Imagens Aéreas de Alta Resolução e o Google Earth” , Agroportal (<http://www.agroportal.pt/a/2005/rbraga.htm>)